ENCYCLOPEDIA of the MIND

Harold Pashler
Editor-in-Chief
ENCYCLOPEDIA of the MIND
Editorial Board

Editor-in-Chief
Harold Pashler
University of California, San Diego

Associate Editor
Emily Goldknopf
University of California, Los Angeles

Editorial Board
Tim Crane
University of Cambridge

Fernanda Ferreira
University of Edinburgh

Marcel Kinsbourne
New School for Social Research

Richard Zemel
University of Toronto

Editorial Assistant
Ruth Hsiao
University of Illinois College of Medicine
## Contents

### Volume 1

List of Entries  vii  
Reader’s Guide  xi  
About the Editor-in-Chief  xix  
Contributors  xxi  
Introduction  xxix  

### Entries

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
<td>F</td>
<td>343</td>
</tr>
<tr>
<td>B</td>
<td>105</td>
<td>G</td>
<td>369</td>
</tr>
<tr>
<td>C</td>
<td>127</td>
<td>H</td>
<td>379</td>
</tr>
<tr>
<td>D</td>
<td>207</td>
<td>I</td>
<td>397</td>
</tr>
<tr>
<td>E</td>
<td>273</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Volume 2

List of Entries  vii  
Reader’s Guide  xi  

### Entries

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>J</td>
<td>441</td>
<td>R</td>
<td>627</td>
</tr>
<tr>
<td>K</td>
<td>447</td>
<td>S</td>
<td>663</td>
</tr>
<tr>
<td>L</td>
<td>455</td>
<td>T</td>
<td>731</td>
</tr>
<tr>
<td>M</td>
<td>477</td>
<td>U</td>
<td>749</td>
</tr>
<tr>
<td>N</td>
<td>553</td>
<td>V</td>
<td>757</td>
</tr>
<tr>
<td>O</td>
<td>569</td>
<td>W</td>
<td>775</td>
</tr>
<tr>
<td>P</td>
<td>577</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Index  797
List of Entries

Access Consciousness
Action and Bodily Movement
Action Slips
Addiction
Affective Forecasting
Aging, Memory, and Information Processing Speed
Allais Paradox
Amnesia
Analogical Mapping and Reasoning
Analogical Reasoning, Models of Development
Anchoring
Anesthesia and Awareness
Anomalous Monism
Anosognosia
Anti-Individualism About Cognition
Anxiety Disorders
Aphasia
Apraxia
Argument Mapping
Atomism About Concepts
Attention, Evolutionary Perspectives
Attention, Neuroimaging Studies of
Attention, Resource Models
Attention and Action
Attention and Consciousness
Attention and Emotion
Attention and Emotions, Computational Perspectives
Attentional Blink Effect
Attitude Change
Attitudes and Behavior
Attraction
Attribution Theory
Audition, Neural Basis
Auditory Masking
Autism
Automatic Behavior
Automaticity
Autoscopic Phenomena
Availability Heuristic
Behavioral Therapy
Behaviorism
Belief and Judgment
Bilingual Language Processing
Bilingualism, Cognitive Benefits of
Black English Vernacular (Ebonics)
Blindsight
Borderline Personality Disorder
Capgras Delusion
Case-Based Reasoning, Computational Perspectives
Categorization, Neural Basis
Categorization, Psychological Perspectives
Category Learning, Computational Perspectives
Causal Theories of Intentionality
Causal Theories of Memory
Change Blindness
Character and Personality, Philosophical Perspectives
Classical Theory of Concepts
Cognitive Dissonance
Cohort Model of Auditory Word Recognition
Collective Action
Common Coding
Compound Words, Processing of
Computational Models of Emotion
Concepts, Comparative Perspectives
Concepts, Development of
Concepts, Philosophical Issues
Concepts and Language
Conceptual Combination
Conduction Aphasia
Confabulation
Conscious Thinking
Consciousness, Comparative Perspectives
Consciousness and Embodiment
Consciousness and the Unconscious
Conversation and Dialogue
<table>
<thead>
<tr>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
<th>Entries</th>
</tr>
</thead>
</table>
Language Development
Language Development, Overregulation in
Language Production, Agreement in
Language Production, Incremental
Processing in
Layered Control Architectures
Learning Styles
Legal Reasoning, Psychological
Perspectives
Lie Detection
Love

Machine Speech Recognition
McCullough Effect
Memory, Interference With
Memory, Neural Basis
Memory and Knowledge
Memory Recall, Dynamics
Mental Action
Mental Causation
Mental Effort
Meta-Analysis
Metacognition and Education
Metaphor
Microgenesis of Consciousness
Mind-Body Problem
Mirror Neurons
Mnemonic Strategies
Modeling Causal Learning
Moral Development
Motivated Thinking
Motor Learning, Practical Aspects
Motor System, Development of
Multimodal Conversational Systems
Multinomial Modeling
Multiple Intelligences Theory
Multitasking and Human Performance
Music and the Evolution of Language
Music Perception

Naïve Realism
Narcissistic Personality Disorder
Natural Action Selection, Modeling
Natural Language Generation
Neural Correlates of Consciousness
Neurodynamics of Visual Search
Neuroeconomics

Object-Dependent Thought
Objects of Memory
Optic Flow

Perceptual Consciousness and Attention
Perceptual Constancy
Personal Identity
Personal Identity, Development of
Personality: Individual Versus Situation Debate
Perspective Taking in Language Processing
Persuasion
Phenomenology of Action
Philosophy of Action
Physicalism
Placebo Effect
Planning in Language Production
Political Psychology
Preconscious Free Will
Prediction, Clinical Versus Actuarial
Production of Language
Prosody in Production
Psychological Refractory Period

Rationality of Emotion
Reaction Time
Realism and Instrumentalism
Reductive Physicalism
Regret
Rehearsal and Memory
Reinforcement Learning, Psychological
Perspectives
Relationships, Development of
Religion and Psychiatry
Representational Theory of Mind
Representations, Development of
Representativeness Heuristic
Resentment
Retrieval Practice (Testing) Effect

Schizophrenia
Scientific Reasoning
Self, Development of
Self-Consciousness
Self-Knowledge
Semantic Dementia
Semantic Memory
Semantic Memory, Computational Perspectives
Sequential Memory, Computational Perspectives
Serial Order Memory, Computational Perspectives
Similarity
Skill Learning, Enhancement of
Sleep and Dreams
Smell, Philosophical Perspectives
<table>
<thead>
<tr>
<th>Social Cognition</th>
<th>Unconscious Emotions, Psychological Perspectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Loafing</td>
<td>Unconscious Perception</td>
</tr>
<tr>
<td>Spacing Effect</td>
<td>Visual Imagery</td>
</tr>
<tr>
<td>Spacing Effect, Practical Applications</td>
<td>Visual Masking</td>
</tr>
<tr>
<td>Spatial Cognition, Development of</td>
<td>Visual Search</td>
</tr>
<tr>
<td>Speech Perception</td>
<td>Visual Working Memory</td>
</tr>
<tr>
<td>Stereopsis</td>
<td>Visuospatial Reasoning</td>
</tr>
<tr>
<td>Stroop Effect</td>
<td>Voluntary Action, Illusion of</td>
</tr>
<tr>
<td>Subliminal Perception</td>
<td>Williams Syndrome</td>
</tr>
<tr>
<td>Synesthesia</td>
<td>Wisdom of Crowds Effect</td>
</tr>
<tr>
<td>Syntactic Production, Agreement in</td>
<td>Word Learning</td>
</tr>
<tr>
<td></td>
<td>Word Recognition, Auditory</td>
</tr>
<tr>
<td>Taste, Philosophical Perspectives</td>
<td>Word Recognition, Visual</td>
</tr>
<tr>
<td>Teleology</td>
<td>Working Memory</td>
</tr>
<tr>
<td>Theory of Appearing</td>
<td>Working Memory, Evolution of</td>
</tr>
<tr>
<td>Thinking</td>
<td>Working Memory in Language Processing</td>
</tr>
<tr>
<td>Time Perception</td>
<td></td>
</tr>
<tr>
<td>Two System Models of Reasoning</td>
<td></td>
</tr>
</tbody>
</table>
Reader’s Guide

Action and Motor Control
Computational Perspectives
Layered Control Architectures
Natural Action Selection, Modeling
Development
Motor System, Development of
Disorders and Pathology
Apraxia
Neural Basis
Apraxia
Common Coding
Desire
Mirror Neurons
Preconscious Free Will
Philosophical Perspectives
Action and Bodily Movement
Collective Action
Desire
Explanation of Action
Freedom of Action
Joint or Collective Intention
Know-How, Philosophical Perspectives
Mental Action
Phenomenology of Action
Philosophy of Action
Preconscious Free Will
Teleology
Psychological Research
Action Slips
Attention and Action
Common Coding
Motor System, Development of
Multitasking and Human Performance
Optic Flow
Preconscious Free Will
Psychological Refractory Period
Voluntary Action, Illusion of

Attention
Computational Perspectives
Attention and Emotions, Computational Perspectives
Evolutionary Perspectives
Attention, Evolutionary Perspectives
Neural Basis
Attention, Evolutionary Perspectives
Attention, Neuroimaging Studies of Inhibition of Return
Mental Effort
Neurodynamics of Visual Search
Philosophical Perspectives
Attention and Consciousness
Psychological Research
Attention, Resource Models
Attention and Action
Attention and Emotion
Attentional Blink Effect
Automaticity
Change Blindness
Divided Attention and Memory
Inattentional Blindness
Inhibition of Return
Mental Effort
Multitasking and Human Performance
Neurodynamics of Visual Search
Perceptual Consciousness and Attention
Psychological Refractory Period
Stroop Effect
Visual Search
Concepts and Categories

Computational Perspectives
Category Learning, Computational Perspectives
Semantic Memory, Computational Perspectives

Development
Concepts, Development of
Representations, Development of
Self, Development of

Disorders and Pathology
Semantic Dementia

Evolutionary Perspectives
Concepts, Comparative Perspectives

Neural Basis
Categorization, Neural Basis

Philosophical Perspectives
Atomism About Concepts
Causal Theories of Intentionality
Classical Theory of Concepts
Concepts, Philosophical Issues
Concepts and Language
Microgenesis of Consciousness
Representational Theory of Mind

Psychological Research
Categorization, Psychological Perspectives
Concepts, Comparative Perspectives
Conceptual Combination
Self, Development of
Semantic Memory
Similarity

Consciousness

Disorders and Pathology
Anosognosia
Automatic Behavior
Autoscopic Phenomena

Evolutionary Perspectives
Consciousness, Comparative Perspectives

Neural Basis
Anesthesia and Awareness
Anosognosia
Microgenesis of Consciousness

Neural Correlates of Consciousness
Preconscious Free Will
Self-Consciousness
Sleep and Dreams

Philosophical Perspectives
Access Consciousness
Attention and Consciousness
Conscious Thinking
Consciousness and Embodiment
Consciousness and the Unconscious
Emotions and Consciousness
Explanatory Gap
Introspection
Microgenesis of Consciousness
Realism and Instrumentalism
Self-Consciousness
Self-Knowledge

Psychological Research
Anesthesia and Awareness
Attention and Consciousness
Consciousness and the Unconscious
Metacognition and Education
Perceptual Consciousness and Attention
Self-Consciousness
Self-Knowledge
Subliminal Perception
Unconscious Emotions, Psychological Perspectives
Unconscious Perception
Voluntary Action, Illusion of

Decision Making

Computational Perspectives
Decision Making and Reward, Computational Perspectives
Natural Action Selection, Modeling

Neural Basis
Decision Making, Neural Underpinnings
Dictator Game
Neuroeconomics

Philosophical Perspectives
Decision Theory, Philosophical Perspectives
Dutch Book Arguments

Practical Applications
Debiasing
Decision Improvement Technologies
Psychological Research
Affective Forecasting
Allais Paradox
Cognitive Dissonance
Debiasing
Dictator Game
Dissent, Effects on Group Decisions
Endowment Effect
Group Decision Making
Wisdom of Crowds Effect

Disorders and Pathology
Addiction
Amnesia
Anosognosia
Anxiety Disorders
Aphasia
Apraxia
Autism
Automatic Behavior
Autoscopic Phenomena
Behavioral Therapy
Borderline Personality Disorder
Capgras Delusion
Conduction Aphasia
Confabulation
Delusions
Dyslexia, Acquired
Dyslexia, Developmental
Dyslexia, Phonological Processing in
Emotion and Psychopathology
Fregoli Delusion
Hypochondria
Narcissistic Personality Disorder
Religion and Psychiatry
Schizophrenia
Semantic Dementia
Williams Syndrome

Emotion
Computational Perspectives
Attention and Emotions, Computational Perspectives
Computational Models of Emotion
Facial Expressions, Computational Perspectives

Cultural Perspectives
Emotion, Cultural Perspectives

Disorders and Pathology
Anxiety Disorders
Religion and Psychiatry

Evolutionary Perspectives
Consciousness, Comparative Perspectives

Neural Basis
Desire
Emotion, Cerebral Lateralization
Emotion, Psychophysiology of
Emotion and Psychopathology
Emotional Recognition, Neuropsychology of

Philosophical Perspectives
Desire
Emotion and Moral Judgment
Emotions and Consciousness
Intentionality of Emotion
Rationality of Emotion

Psychological Research
Affective Forecasting
Anxiety Disorders
Attention and Emotion
Cognitive Dissonance
Disgust
Embarrassment
Emotion, Psychophysiology of
Emotion, Structural Approaches
Emotion and Moral Judgment
Emotion and Working Memory
Emotion Regulation
Envy
Facial Expressions, Emotional
Guilt
Happiness
Jealousy
Love
Regret
Religion and Psychiatry
Resentment
Unconscious Emotions, Psychological Perspectives

Epistemology
Philosophical Perspectives
Anti-Individualism About Cognition
Belief and Judgment
Causal Theories of Memory
Conscious Thinking
Consciousness and the Unconscious
Deductive Reasoning
Descriptive Thought
Emotion and Moral Judgment
Experimental Philosophy
Explanatory Gap
Introspection
Know-How, Philosophical Perspectives
Knowledge by Acquaintance
Memory and Knowledge
Self-Knowledge
Psychological Research
Self-Knowledge

Genetics
Heritability

Disorders and Pathology
Addiction
Schizophrenia
Williams Syndrome

Evolutionary Perspectives
Genes and Linguistic Tone

Groups
Cultural Perspectives
Distributed Cognition

Philosophical Perspectives
Collective Action
Joint or Collective Intention

Psychological Research
Dissent, Effects on Group Decisions
Distributed Cognition
Group Decision Making
Intergroup Conflict
Intergroup Conflict, Models of

Intelligence
Neural Basis
Intelligence, Neural Basis

Psychological Research
Flynn Effect: Rising Intelligence Scores
Heritability
Intelligence and Working Memory
Multiple Intelligences Theory

Language and Communication

Computational Perspectives
Machine Speech Recognition
Multimodal Conversational Systems
Natural Language Generation

Cultural Perspectives
Bilingual Language Processing
Bilingualism, Cognitive Benefits of
Black English Vernacular (Ebonics)
Gender Differences in Language and Language Use
Gesture and Language Processing
Heritage Language and Second Language Learning

Development
Dyslexia, Developmental
Dyslexia, Phonological Processing in
Heritage Language and Second Language Learning
Innateness and Parameter Setting
Language Development
Language Development, Overregulation in
Word Learning

Disorders and Pathology
Aphasia
Conduction Aphasia
Dyslexia, Acquired
Dyslexia, Developmental
Dyslexia, Phonological Processing in
Semantic Dementia

Evolutionary Perspectives
Genes and Linguistic Tone
Music and the Evolution of Language

Neural Basis
Aphasia
Conduction Aphasia
Philosophical Perspectives
Descriptions
Descriptive Thought
Indexical Thought
Intension and Extension
Metaphor
Object-Dependent Thought
Practical Applications
Deception, Linguistic Cues to
Lie Detection
Psychological Research
Bilingual Language Processing
Bilingualism, Cognitive Benefits of
Cohort Model of Auditory Word Recognition
Compound Words, Processing of
Concepts and Language
Conceptual Combination
Conversation and Dialogue
Declarative/Procedural Model of Language
Discourse Processing, Models of
Disfluencies: Comprehension Processes
Eye Movements During Reading
Frequency Effects in Word Recognition
Gesture and Language Processing
Inferences in Language Comprehension
Inner Speech
Language Development
Language Production, Agreement in
Language Production, Incremental
Processing in
Lie Detection
Metaphor
Perspective Taking in Language Processing
Planning in Language Production
Production of Language
Prosody in Production
Speech Perception
Syntactic Production, Agreement in
Word Learning
Word Recognition, Auditory
Word Recognition, Visual
Working Memory in Language Processing

Learning
Computational Perspectives
Modeling Causal Learning

Development
Knowledge Acquisition in Development
Word Learning
Neural Basis
Desire
Exercise and the Brain
Human Classification Learning
Practical Applications
Behavioral Therapy
Desirable Difficulties Perspective on Learning
Discrimination Learning, Training Methods
Exercise and the Brain
Learning Styles
Metacognition and Education
Motor Learning, Practical Aspects
Placebo Effect
Retrieval Practice (Testing) Effect
Skill Learning, Enhancement of
Spacing Effect, Practical Applications
Psychological Research
Behavioral Therapy
Behaviorism
Desirable Difficulties Perspective on Learning
Desire
Human Classification Learning
Learning Styles
Metacognition and Education
Placebo Effect
Reinforcement Learning, Psychological
Perspectives
Retrieval Practice (Testing) Effect
Skill Learning, Enhancement of
Spacing Effect
Word Learning

Memory
Computational Perspectives
Semantic Memory, Computational Perspectives
Sequential Memory, Computational Perspectives
Serial Order Memory, Computational Perspectives

Development
Aging, Memory, and Information Processing Speed
Event Memory, Development
Disorders and Pathology
Amnesia
Confabulation
Semantic Dementia

Evolutionary Perspectives
Working Memory, Evolution of

Neural Basis
Amnesia
Event Memory, Development
Implicit Memory
Memory, Neural Basis

Philosophical Perspectives
Causal Theories of Memory
Memory and Knowledge
Objects of Memory

Practical Applications
Exercise and the Brain
Retrieval Practice (Testing) Effect
Mnemonic Strategies
Spacing Effect, Practical Applications

Psychological Research
Aging, Memory, and Information Processing
  Speed
Divided Attention and Memory
Emotion and Working Memory
Event Memory, Development
Exercise and the Brain
Eyewitness Memory
Implicit Memory
Intelligence and Working Memory
Memory, Interference With
Memory Recall, Dynamics
Rehearsal and Memory
Retrieval Practice (Testing) Effect
Semantic Memory
Spacing Effect
Visual Working Memory
Working Memory
Working Memory in Language Processing

Metaphysics
Philosophical Perspectives
Anomalous Monism
Anti-Individualism About Cognition

Causal Theories of Intentionality
Conscious Thinking
Consciousness and Embodiment
Eliminative Materialism
Emergence
Extended Mind
Freedom of Action
Idealism
Intentionality of Bodily Sensation
Intentionality of Emotion
Mental Action
Mental Causation
Naïve Realism
Object-Dependent Thought
Personal Identity
Philosophy of Action
Physicalism
Realism and Instrumentalism
Reductive Physicalism
Smell, Philosophical Perspectives
Taste, Philosophical Perspectives
Teleology
Theory of Appearing

Methodology
Neural Basis
Electrophysiological Studies of Mind

Philosophical Perspectives
Experimental Philosophy

Practical Applications
Prediction, Clinical Versus Actuarial

Psychological Research
Heritability
Meta-Analysis
Multinomial Modeling
Reaction Time

Mind-Body Problem
Philosophical Perspectives
Action and Bodily Movement
Anomalous Monism
Consciousness and Embodiment
Eliminative Materialism
Emergence
Explanatory Gap
Idealism
Know-How, Philosophical Perspectives
Mental Causation
Mind-Body Problem
Physicalism
Realism and Instrumentalism
Reductive Physicalism
Representational Theory of Mind
Teleology

Morality/Ethics

Development
Moral Development

Philosophical Perspectives
Character and Personality, Philosophical Perspectives
Emotion and Moral Judgment
Joint or Collective Intention
Philosophy of Action

Psychological Research
Emotion and Moral Judgment
Guilt
Moral Development
Political Psychology

Perception

Computational Perspectives
Face Recognition in Humans and Computers
Facial Expressions, Computational Perspectives

Neural Basis
Audition, Neural Basis
Blindsight
Common Coding
Emotional Recognition, Neuropsychology of Mirror Neurons
Synesthesia
Visual Imagery

Philosophical Perspectives
Disjunctive Theory of Perception
Hearing, Philosophical Perspectives
Naïve Realism
Smell, Philosophical Perspectives
Taste, Philosophical Perspectives

Psychological Research
Auditory Masking
Common Coding
Depth Perception
Face Perception
Facial Expressions, Emotional
Hearing
McCollough Effect
Music Perception
Optic Flow
Perceptual Consciousness and Attention
Perceptual Constancy
Speech Perception
Stereopsis
Subliminal Perception
Synesthesia
Time Perception
Unconscious Perception
Visual Imagery
Visual Masking
Visual Search
Visuospatial Reasoning
Word Recognition, Auditory
Word Recognition, Visual

Personality

Development
Personal Identity, Development of

Disorders and Pathology
Borderline Personality Disorder
Narcissistic Personality Disorder

Philosophical Perspectives
Character and Personality, Philosophical Perspectives

Psychological Research
Attitude Change
Borderline Personality Disorder
Personal Identity, Development of
Personality: Individual Versus Situation: Debate
Political Psychology

The Social Mind

Development
Relationships, Development of
Disorders and Pathology
Capgras Delusion

Philosophical Perspectives
Folk Psychology

Psychological Research
Attitude Change
Attitudes and Behavior
Attraction
Attribution Theory
Cognitive Dissonance
Group Decision Making
Intergroup Conflict
Intergroup Conflict, Models of
Love
Personality: Individual Versus Situation
Debate
Persuasion
Political Psychology
Social Cognition
Social Loafing
Wisdom of Crowds Effect

Thinking and Reasoning
Computational Perspectives
Analogical Reasoning, Models of Development
Case-Based Reasoning, Computational Perspectives

Cultural Perspectives
Distributed Cognition

Development
Analogical Reasoning, Models of Development
Knowledge Acquisition in Development
Spatial Cognition, Development of

Philosophical Perspectives
Deductive Reasoning

Practical Applications
Argument Mapping
Case-Based Reasoning, Computational Perspectives

Psychological Research
Analogical Mapping and Reasoning
Anchoring
Availability Heuristic
Distributed Cognition
Legal Reasoning, Psychological Perspectives
Motivated Thinking
Representativeness Heuristic
Scientific Reasoning
Spatial Cognition, Development of Thinking
Two System Models of Reasoning
Visual Imagery
Visuospatial Reasoning
About the Editor-in-Chief

Harold Pashler is Distinguished Professor of Psychology at the University of California, San Diego, where he heads the Learning, Attention, and Perception Lab and serves on the faculty of the Cognitive Science Interdisciplinary Program. In 1999, Pashler received the Troland Award from the National Academy of Sciences. The Academy cited Pashler’s “many experimental breakthroughs in the study of spatial attention and executive control, and . . . his insightful analysis of human cognitive architecture.” He is also a Fellow of the Society of Experimental Psychologists (SEP) and of the Association for Psychological Science (APS).

Pashler authored and edited two highly cited classics in the field of attention (*The Psychology of Attention* and *Attention*—both published in 1998). He was also the general editor of the third edition of the four-volume, influential *Stevens Handbook of Experimental Psychology*.

Pashler’s research accomplishments span many areas of cognitive psychology. In 1988 he was the first to report a phenomenon that has come to be called change blindness (an observer’s inability to detect even major changes in displays that flicker off and on again while the observer is staring intently at the display). A large literature has sprung up that explores the philosophical and psychological implications of this phenomenon.

Pashler’s interest in the nature of visual awareness also helped spark the development (jointly with Liqiang Huang, now at the Chinese University of Hong Kong) of the Boolean maps theory of visual attention, published in *Science* and *Psychological Review*. The theory seeks to characterize the logical limitations of human visual awareness.

Pashler’s most sustained research efforts, however, have focused on a different aspect of attention. In a long series of influential papers published in the 1980s and 1990s, he and his students and collaborators developed “chronometric” (reaction-time based) techniques for analyzing the attentional bottlenecks that arise when people try to perform different tasks at the same time. Pashler showed that different bottleneck models make distinctive and precise predictions for the effects of experimental manipulations targeted at different stages of multiple tasks being performed concurrently. The results indicated that the brain’s key processing bottleneck is in decision making and the selection of responses (whereas other mental operations like perception and actually executing motor actions are not subject to the bottleneck). He also conducted research using driving simulators to show that these findings generalize to cases where drivers engage in unrelated tasks while following another car that slows from time to time (here the bottleneck resulted in delays in the application of the brakes).

In the past few years, Pashler has embarked on a vigorous analysis of the factors that affect the efficiency of human learning, seeking to find concrete methods of instruction and training that enhance the efficiency and durability of learning. The U.S. Department of Education commissioned Pashler to chair an expert panel that produced a practice guide for educators, textbook authors, and educational software developers (*Organizing Instruction and Study to Improve Student Learning*, published in 2007). The practice guide and an accompanying website have been consulted by many thousands of educators. In 2009, Pashler also chaired a review commissioned by *Psychological Science in the Public Interest* that drew considerable press attention for its highly critical conclusions about some commonly accepted ideas about learning styles.

In recent years, Pashler has collaborated with researchers from machine learning (chiefly Michael Mozer from the University of Colorado Department of Computer Science) to explore the neuro-computational basis of learning enhancement. Finally, in collaboration with Edward Vul, Pashler uncovered
what has come to be called the *wisdom of inner crowds effect* in the field of judgment and decision making—helping trigger research efforts in multiple centers to develop new strategies for enhancing real-world skills in estimation and other tasks requiring accurate judgment.

Pashler has also served as an associate editor of *Psychonomic Bulletin & Review* and is on the editorial board of the *Journal of Experimental Psychology*, the *American Journal of Psychology*, *Visual Cognition*, and other major journals in the field. Together with several colleagues, in 2012 he co-founded the website PsychFileDrawer.org to promote replication attempts in psychology and to combat the “file drawer problem” that has drawn increasing concern in recent years.

Pashler received undergraduate degrees in psychology and in logic and philosophy of science with honors from Brown University and a PhD in psychology from the University of Pennsylvania (1985).
Contributors

Phillip L. Ackerman  
*Georgia Institute of Technology*

Karen E. Adolph  
*New York University*

Alan Allport  
*Oxford University*

Nancy Alvarado  
*Columbia State Polytechnic University, Pomona*

Maria Alvarez  
*King's College London*

Julio Arboleda-Flórez  
*Queen's University*

Arthur Aron  
*State University of New York at Stony Brook*

F. Gregory Ashby  
*University of California, Santa Barbara*

Peter Bachman  
*University of California, Los Angeles*

Linden J. Ball  
*Lancaster University*

Dale J. Barr  
*University of California, Riverside*

Lisa Feldman Barrett  
*Northeastern University*

William H. Batchelder  
*University of California, Irvine*

Clare Batty  
*University of Kentucky*

Patricia J. Bauer  
*Emory University*

John Baugh  
*Washington University in St. Louis*

Carrie E. Bearden  
*University of California, Los Angeles*

Jacob Scott Beck  
*York University in Toronto*

Mark W. Becker  
*Micigan State University*

Christian Beenfeldt  
*University of Copenhagen*

Sarah E. Berger  
*College of Staten Island, City University of New York*

Derek Besner  
*University of Waterloo*

Michael A. Bishop  
*Florida State University*

Robert Allen Bjork  
*University of California, Los Angeles*

Karen Blackmon  
*New York University*

Galen V. Bodenhausen  
*Northwestern University*

Joan C. Borod  
*Queens College, City University of New York*

Guy Bosmans  
*KU Leuven*

Matthew M. Botvinick  
*Princeton University*

Thomas J. Bouchard Jr.  
*University of Minnesota*

Theo K. Bouman  
*University of Groningen*

Howard Bowman  
*University of Kent at Canterbury*

Darren Bradley  
*City College of New York*

Caroline Braet  
*Ghent University*

Bruno G. Breitmeyer  
*University of Houston*

Susan E. Brennan  
*Stony Brook University*

Berit Brogaard  
*University of Missouri, St. Louis*

Tobias Brosch  
*University of Geneva*

Jason Walter Brown  
*New York University Medical Center*

Peter Brugger  
*University Hospital Zurich*

Joanna J. Bryson  
*University of Bath*

Andrew C. Butler  
*Duke University*

Stephen Andrew Butterfill  
*University of Warwick, UK*

Mark John Cain  
*Oxford Brookes University*
William Keith Campbell  
*University of Georgia*

David Caplan  
*Harvard Medical School*

Daniel Casasanto  
*New School for Social Research*

Nicholas J. Cepeda  
*York University*

Joyce Chai  
*Michigan State University*

Yulia Chentsova Dutton  
*Georgetown University*

Ana B. Chica  
*Centre de Recherche de l’Institut du Cerveau et de la Moelle Epinière*

Colby Chlebowski  
*University of Connecticut*

Nicholas J. S. Christenfeld  
*University of California, San Diego*

Paul M. Churchland  
*University of California, San Diego*

Jonathan D. Cohen  
*Princeton University*

Sam Coleman  
*University of Hertfordshire*

Annalisa Coliva  
*University of Modena and Reggio Emilia*

Max Coltheart  
*Macquarie University*

Frederick L. Coolidge  
*University of Colorado, Colorado Springs*

Joel M. Cooper  
*University of Utah*

Richard P. Cooper  
*Birkbeck, University of London*

Randall C. Cork  
*Alaska VAHSRO*

Martin Corley  
*University of Edinburgh*

Kelly E. Courtney  
*University of California, Los Angeles*

Nelson Cowan  
*University of Missouri-Columbia*

Alan Cowey  
*University of Oxford*

Stephen Crain  
*Macquarie University*

Sean Crawford  
*University of Manchester*

Valentina Daelli  
*International School for Advanced Studies*

Delphine Dahan  
*University of Pennsylvania*

Ryan S. Darby  
*University of California, San Diego*

Adam Darlow  
*Brown University*

Eddy J. Davelaar  
*Birkbeck, University of London*

Matthew H. Davis  
*Medical Research Council Cognition and Brain Sciences Unit*

Samuel B. Day  
*Susquehanna University*

Ronald de Sousa  
*University of Toronto*

Gedeon Deák  
*University of California, San Diego*

Gustavo Deco  
*Institució Catalana de Recerca i Estudis Avançats/University Pompeu Fabra*

Dan Dediu  
*Max Planck Institute for Psycholinguistics*

Sergio Della Sala  
*University of Edinburgh*

Simon Dennis  
*The Ohio State University*

Michaela Dewar  
*University of Edinburgh*

John W. Donahoe  
*University of Massachusetts, Amherst*

Kristie R. Dukewich  
*Dalhousie University*

Kevin N. Dunbar  
*University of Toronto Scarborough*

John Dunlosky  
*Kent State University*

Elizabeth W. Dunn  
*University of British Columbia*

Sorin Dusan  
*MCT Inc./NASA Ames Research Center*

Kenny Easwaran  
*University of Southern California*

Phoebe C. Ellsworth  
*University of Michigan*

Nadine Elzein  
*University College London*

Nicholas Epley  
*University of Chicago*

Kirk Erickson  
*University of Pittsburgh*

Leandre R. Fabrigar  
*Queen’s University*

Simon Farrell  
*University of Bristol*

Deborah Fein  
*University of Connecticut*
Myra Annette Fernandes
University of Waterloo

Jordi Fernández
University of Adelaide

Charles Fernyhough
Durham University

Victor S. Ferreira
University of California, San Diego

William Fish
Massey University

W. Tecumseh Fitch
University of Vienna

Evelyn A. Forster
University of Toronto, Scarborough

Claudette Fortin
Université Laval

Julie Franck
University of Geneva

Vittorio Gallese
University of Parma

Brian Garrett
Australian National University

Susan A. Gelman
University of Michigan

Margaret Gilbert
University of California, Irvine

Andrew Glenster
University of Reading

Paul W. Glimcher
New York University

Philip Goff
University of Liverpool

Sanford Goldberg
Northwestern University

Peter Goldie
University of Manchester

Robert L. Goldstone
Indiana University

Tamar H. Gollan
University of California, San Diego

Usha Goswami
University of Cambridge

Zenzi M. Griffin
University of Texas at Austin

James J. Gross
Stanford University

Rosanna E. Guadagno
University of Alabama

Sören Häggqvist
Stockholm University

Richard J. Haier
University of California, Irvine

Eric Halgren
University of California, San Diego

Eddie Harmon-Jones
Texas A&M University

Christine R. Harris
University of California, San Diego

Susan Harter
University of Denver

Alice F. Healy
University of Colorado, Boulder

Gregory Hickok
University of California, Irvine

E. Tory Higgins
Columbia University

William Hirstein
Elmhurst College

Merrill Hiscock
University of Houston

Christoph Hoerl
University of Warwick, UK

Jakob Hohwy
Monash University

Joseph Hopfinger
University of North Carolina

Xiangen Hu
University of Memphis

Edward Michael Hubbard
University of Wisconsin, Madison

Scott A. Huettel
Duke University

Daniel D. Hutto
University of Hertfordshire

William G. Iacono
University of Minnesota

Yoel Inbar
Tilburg University

Maria Jalbrzikowski
University of California, Los Angeles

Linda Jeffery
University of Western Australia

Erin L. Jones
University of Illinois

Steve Joordens
University of Toronto Scarborough

J. Scott Jordan
Illinois State University

Peter Juslin
Uppsala University

Michael Kahana
University of Pennsylvania

Mariia Kaliuzhna
École Polytechnique Fédérale de Lausanne

Irene P. Kan
Villanova University

Annette Karmiloff-Smith
Birkbeck, University of London

Timothy Ketelaar
New Mexico State University
Boaz Keysar  
*University of Chicago*

John F. Kihlstrom  
*University of California, Berkeley*

Irving Kirsch  
*University of Hull*

Suzanne C. Klenck  
*University of Houston*

Barbara Knowlton  
*University of California, Los Angeles*

Harold G. Koenig  
*Duke University*

Nate Kornell  
*Williams College*

Daniel C. Krawczyk  
*University of Texas at Dallas and UT Southwestern*

Maia S. Kredentser  
*Queen’s University*

Thomas Kroedel  
*University of Konstanz*

Jane Kroger  
*University of Tromsoe*

Victor Kuperman  
*McMaster University*

D. Robert Ladd  
*University of Edinburgh*

John E. Laird  
*University of Michigan*

Harold Langsam  
*University of Virginia*

Marina Larkina  
*Emory University*

Richard P. Larrick  
*Duke University*

David Leake  
*Indiana University*

Amy E. Learmonth  
*William Patterson University*

Kathryn A. Leech  
*University of Maryland*

Hayley C. Leonard  
*Goldsmiths, University of London*

Jacqueline S. Leventon  
*Emory University*

Joseph Levine  
*University of Massachusetts, Amherst*

Daniel J. Levitin  
*McGill University*

Ifat Levy  
*Yale University*

Gary W. Lewandowski Jr.  
*Monmouth University*

Stephan Lewandowsky  
*University of Western Australia*

Rebecca Lewthwaite  
*Rancho Los Amigos National Rehabilitation Center*

Li Li  
*University of Hong Kong*

Debra L. Long  
*University of California, Davis*

Guy Longworth  
*University of Warwick, UK*

Dan Lopez-Paniagua  
*Colorado State University*

Bradley C. Love  
*University College London/University of Texas at Austin*

Hongjing Lu  
*University of California, Los Angeles*

Roy Lubit  
*Forensic Psychiatric Consultants*

Margaret T. Lynn  
*University of Ghent*

Edouard Machery  
*University of Pittsburgh*

Arien Mack  
*New School for Social Research*

Colin M. MacLeod  
*University of Waterloo*

Bruce Duncan MacQueen  
*University of Gdansk*

James S. Magnuson  
*University of Connecticut and Haskins Laboratories*

Svein Magnussen  
*University of Oslo*

Heidi L. Maibom  
*Carleton University*

John C. Malone  
*University of Tennessee, Knoxville*

Jean M. Mandler  
*University of California, San Diego*

Albert E. Mannes  
*University of Pennsylvania*

Neil C. Manson  
*Lancaster University*

Anthony Marcel  
*University of Hertfordshire*

Maria M. Margariti  
*University of Athens*

Robert P. Marinier III  
*Soar Technology, Inc.*

Clark McCauley  
*Bryn Mawr College*

Kristina L. McDonald  
*University of Alabama*

Joseph T. McGuire  
*University of Pennsylvania*

Ryan McKay  
*Royal Holloway, University of London*

Kateri McRae  
*University of Denver*
Nathan Medeiros-Ward
University of Utah

Darya Melnyk
University of Florida

Erica P. Meltzer
Queens College, City University of New York

Richard Menary
Macquarie University

Miriam Meyerhoff
University of Auckland

Joseph A. Mikels
DePaul University

Jonathan Miller
University of Pennsylvania

Joshua David Miller
University of Georgia

David B. Mitchell
Kennesaw State University

Andreas Mojzisch
Georg-August-University Goettingen

Daniel C. Molden
Northwestern University

Christopher Mole
University of British Columbia

Barbara Montero
City University of New York

Silvina Montrul
University of Illinois at Urbana-Chamiga

Brian C. J. Moore
University of Cambridge

Peter Morris
Lancaster University, UK

Robert G. Morrison
Loyola University Chicago

Ezequiel Morsella
San Francisco State University/
University of California, San Francisco

Morris Moscovitch
University of Toronto

Steven B. Most
University of Delaware

Neil W. Mulligan
University of North Carolina

Nicole L. Muscanell
University of Alabama

Bence Nanay
University of Antwerp

David Navon
University of Haifa

Matthew L. Newman
Arizona State University

Daniel Nolan
Australian National University

Paul Noordhof
University of York

Georg Northoff
University of Ottawa Institute of Mental Health Research

Peter J. Norton
University of Houston

Casey O’Callaghan
Rice University

Haluk Öğmen
University of Houston

Christian N. L. Olivers
Vrije Universiteit Amsterdam

Elizabeth A. Olson
University of Wisconsin, Whitewater

Gary Ostertag
Nassau Community College

Alice J. O’Toole
University of Texas, Dallas

Maria Pachalska
Krakow University

Jaak Panksepp
Washington State University

Philip I. Pavlik Jr.
University of Memphis

Mark Perry
Brunel University

Carissa L. Philippi
University of Wisconsin

Ian B. Phillips
All Souls College, Oxford University

Lawrence H. Pick
Gallaudet University

David Pitt
California State University, Los Angeles

David A. Pizarro
Cornell University

Christopher J. Plack
University of Manchester

Daniel J. Power
University of Northern Iowa

Tony J. Prescott
University of Sheffield

Jesse Prinz
City University of New York, Graduate Center

Wolfgang Prinz
Max Planck Institute for Human Cognition and Brain Science

Simon Prosser
University of Saint Andrews

Michael J. Proulx
University of Bath

Kane Pryor
Weill Medical College of Cornell University

Jordi Quoidbach
Harvard University

Panu Raatikainen
University of Helsinki

Patrick Michael Rabbitt
University of Oxford
Franck Ramus  
Centre National de la Recherche Scientifique

David N. Rapp  
Northwestern University

Kathleen Rastle  
Royal Holloway, University of London

Lara A. Ray  
University of California, Los Angeles

Keith Rayner  
University of California, San Diego

Gregg H. Recanzone  
University of California, Davis

Daniel Reisberg  
Reed College

Ehud Reiter  
University of Aberdeen

Patricia A. Reuter-Lorenz  
University of Michigan

Gillian Rhodes  
University of Western Australia

Marjorie Rhodes  
New York University

Travis A. Riddle  
Columbia University

David M. Riefer  
California State University, San Bernardino

Evan F. Risko  
University of Memphis

Henry L. Roediger III  
Washington University in St. Louis

James C. Root  
Weill Medical College of Cornell University

David Rose  
Carnegie Mellon University

Robert Rosenthal  
University of California, Riverside

Brian H. Ross  
University of Illinois

Meredith L. Rowe  
University of Maryland

Kenneth H. Rubin  
University of Maryland

David Rudrauf  
University of Iowa

Raffaella Ida Rumiati  
International School for Advanced Studies

Robert Douglas Rupert  
University of Colorado, Boulder

Eric Ruthruff  
University of New Mexico

Arthur G. Samuel  
State University of New York at Stony Brook

Lawrence J. Sanna  
University of Michigan

Joshua Schechter  
Brown University

Michael Schmitz  
Universität Konstanz

Simone Schnall  
University of Cambridge

Elizabeth R. Schotter  
University of California, San Diego

Timothy Schroeder  
The Ohio State University

Stefan Schulz-Hardt  
Georg-August-University Goettingen

Aaron Schurger  
NeuroSpin, CEA-Saclay

David Ryan Schurtz  
Stevenson University

Eric Schwitzgebel  
University of California, Riverside

Carol Seger  
Colorado State University

Scott Sehon  
Bowdoin College

Anil Seth  
University of Sussex

James A. Shepperd  
University of Florida

Bhavin R. Sheth  
University of Houston

Renita Silva  
University College London

Patrick Simen  
Oberlin College

Murray Singer  
University of Manitoba

Steven Sloman  
Brown University

Barry C. Smith  
University of London

Richard H. Smith  
University of Kentucky

Declan Smithies  
The Ohio State University

Julie Snowden  
University of Manchester

Jack B. Soll  
Duke University

Mark Solms  
University of Cape Town

Matthew Soteriou  
University of Warwick, UK

Mark Sprevak  
University of Edinburgh

Angela Agnes Stanton  
Claremont Graduate University

Karin Sternberg  
Sternberg Consulting
Robert J. Sternberg  
Tufts University

Jennifer A. Stolz  
University of Waterloo

Debi Stransky  
York University, Centre for Vision Research

David L. Strayer  
University of Utah

Károly Takács  
Corvinus University of Budapest

Eric G. Taylor  
University of Illinois

Keith W. Thiede  
Boise State University

Rosalind Thornton  
Macquarie University

Anna K. Tirovolas  
McGill University

Cain Samuel Todd  
Lancaster University

Natasha Tokowicz  
University of Pittsburgh

David Trafimow  
New Mexico State University

Matthew J. Traxler  
University of California, Davis

Alessandro Treves  
International School for Advanced Studies

J. D. Trout  
Loyola University Chicago

Raimo Tuomela  
University of Helsinki

Elliot Turiel  
University of California, Berkeley

Barbara Tversky  
Teachers College, Columbia University/Stanford University

Katherine E. Tyson  
University of Connecticut

Michael T. Ullman  
Georgetown University

Leaf Van Boven  
University of Colorado, Boulder

Tim van Gelder  
University of Melbourne

Trisha Van Zandt  
The Ohio State University

Ruut Veenhoven  
Erasmus University Rotterdam

Max Velmans  
Goldsmiths, University of London

Mieke Verfaellie  
Memory Disorders Research Center

Jennifer Vonk  
Oakland University

Patrik Olivier Vuilleumier  
University of Geneva

Mark Wagner  
Wagner College

Robert Ward  
Bangor University

Elizabeth A. Ware  
Viterbo University

Lynn Waterhouse  
The College of New Jersey

Michael Watkins  
Auburn University

Duane G. Watson  
University of Illinois at Urbana-Champaign

Catherine Wearing  
Wellesley College

Bethany Weber  
Iowa State University

Renate Wehrle  
Max Planck Institute of Psychiatry, Munich

Gary L. Wells  
Iowa State University

Jacob Westfall  
University of Colorado, Boulder

Linda R. Wheeldon  
University of Birmingham, UK

Laurie M. Wilcox  
York University, Centre for Vision Research

Kenneth Williford  
University of Texas at Arlington

Daniel T. Willingham  
University of Virginia

Piotr Winkielman  
University of California, San Diego

Edward Wisniewski  
University of North Carolina, Greensboro

Jing Wu  
Soochow University, China

Gabriele Wulf  
University of Nevada, Las Vegas

Thomas Wynn  
University of Colorado, Colorado Springs

David Yates  
King’s College London

Garry Young  
Nottingham Trent University

Marcel Zeelenberg  
Tilburg University
Introduction

The study of the mind, at least in the West, was begun by philosophers several thousand years ago and to this day continues to be a major focus of work in philosophy. Over the past several centuries, however, this study has come to occupy an increasing amount of human intellectual activity in many disciplines and to attract increasing public attention. In the 19th century, the discipline of experimental psychology emerged in Germany with an initial focus on memory. In the 20th century, the study of mind became one of the most important (and, in some circles, fashionable) intellectual pursuits of the era as ambitious (some would say, overweening) movements stepped forward with claims of understanding the fundamental nature of the human psyche: psychoanalysis (which emerged from psychiatry) and behaviorism (which emerged from academic experimental psychology).

Near the end of the 20th century, psychologists’ theoretical ambitions retreated as the field became devoted to more highly focused empirical and analytical research. New ventures addressing the nature of thought (cognitive science) and the links between the mind and the brain (cognitive neuroscience) gathered force, propelled by the emergence of computing technologies as well as by new methods for investigating brain activity. As we enter the 21st century, interest in mental life continues to grow across many disciplines, with computational approaches attracting ever greater interest, powered by the successes of contemporary computer scientists in producing striking feats of intelligence (or at least the appearance of intelligence) through reliance upon large-scale statistical computation.

Scope

The Encyclopedia of the Mind is unique in its breadth and scope, providing brief but authoritative entries on topics ranging across the entire landscape just described. The encyclopedia covers topics and perspectives from all the major contemporary disciplines concerned with the study of the mind. Not surprisingly, psychology—often defined as the study of the mind and behavior—looms large here. The coverage of psychology in the encyclopedia includes not only cognitive psychology but also other major areas such as social psychology, developmental psychology, cognitive neuropsychology (the study of how brain damage affects cognitive processes), and the psychology of perception. The psychology of language (including psycholinguistics) and the field of linguistics, insofar as it bears on the mind generally, are also well represented. Philosophy continues to play a pivotal role in this encyclopedia, reflecting the fascinating and often provocative scholarship taking place in the philosophy of mind and metaphysics, with entries on topics such as the mind-body problem, personal identity, and the relationship between human action and underlying physical states.

A unique feature of the encyclopedia is its coverage of computational perspectives on the mind. Beginning in the 1960s, it became clear that a full understanding of mental life would need to include a mechanistic understanding of the kinds of information processing operations that allow intelligent behavior to emerge. The field of artificial intelligence thus bloomed, focused on building artificial minds. In its initial efforts, artificial intelligence focused on high-level symbolic processing approaches. However, in the past 10 years or so, much activity in this area shifted to the banner of “machine learning,” a discipline with strong connections to statistics. Computational approaches are well covered in the Encyclopedia of the Mind.

The editors of the encyclopedia decided at the outset that, although the volumes would cover the brain bases of mental activity to the extent that this
work sheds light on the mind, the Encyclopedia of the Mind would not aspire to be a reference work on the biology of the brain per se.

Organization

The Encyclopedia of the Mind contains 293 entries that fall under the following broad topics: Action and Motor Control, Attention, Concepts and Categories, Consciousness, Decision Making, Disorders and Pathology, Emotion, Epistemology, Genetics, Groups, Intelligence, Language and Communication, Learning, Memory, Metaphysics, Methodology, Mind/Body Problem, Morality/Ethics, Perception, Personality, The Social Mind, and Thinking and Reasoning.

Under each of these broad topics, there are a number of general approaches that recur, including Computational Perspectives, Cultural Perspectives, Development, Evolutionary Perspectives, Neural Basis, Philosophical Perspectives, Practical Applications, and Psychological Research.

This structure—broad topics with cross-cutting research approaches—is reflected in the Reader’s Guide, which provides a convenient way for readers to find entries on a particular topic. Reflecting the interdisciplinary nature of the encyclopedia, many entries discuss more than one topic or discuss topics from more than one perspective. In addition, each entry is followed by cross-references to other entries in the encyclopedia.

How the Encyclopedia Was Created

The preparation of this encyclopedia (which began in 2007) has been a long process involving a great deal of work by many people. There were five basic steps to its creation:

Step 1. In consultation with SAGE, the editor-in-chief appointed four highly distinguished scholars to the editorial board. The board consisted of professors Fernanda Ferreira (University of Edinburgh), an internationally recognized expert in the psychology of language and linguistics; Richard Zemel (University of Toronto Department of Computer Science), an expert in artificial intelligence and machine learning; Marcel Kinsbourne (New School for Social Research), a world-renowned neurologist and neuropsychologist; and Tim Crane (University of Cambridge), one of the most distinguished contemporary philosophers of mind.

Step 2. The editor-in-chief, together with the editorial board, met and decided upon the basic structure for the encyclopedia, as discussed above and as reflected in the Reader’s Guide.

Step 3. Working together, the editorial team identified potential contributors to write entries and explored the availability of these contributors. Many of the suggestions for contributors came from the editorial board, but others emerged from a literature review conducted by the editor-in-chief in consultation with colleagues around the globe. The focus was on identifying potential authors with a broad command of the subject matter and its context. The entry authors include many up-and-coming junior scholars (typically assistant professors or equivalent) but also many highly respected senior figures.

Step 4. When the first draft of each entry arrived, it was reviewed and suggestions for revisions were made. The typical entry was revised and then accepted. Associate Editor Emily Goldknopf, PhD, was crucially involved in this editorial review, under the guidance of the editor-in-chief.

Step 5. In the final step of the process, the encyclopedia underwent further editorial review and copy editing by SAGE Publications, and additional rounds of review by the editors and SAGE proofreaders prior to publication.

Intended Audience

Our goal in commissioning and editing articles for this work was to emphasize accessibility for an educated readership. Authors were instructed to make their work understandable to college and university students in cognitive science and related fields such as philosophy, psychology, and neuroscience, and mathematical formulas were kept to a minimum.

Acknowledgments

A number of individuals played key roles in the creation of the Encyclopedia of the Mind. Jim Brace-Thompson of SAGE Publications helped formulate the concept of the encyclopedia in early discussions with the editor-in-chief. Diana Axelsen of SAGE was
the lead developmental editor for the great majority of the development process. She provided advice and encouragement at all stages along the way. Also at SAGE, Jane Haenel played an important role as production editor.

Working closely with the editor-in-chief at the University of California, San Diego, Editorial Assistant Ruth Hsiao (currently a medical student at the University of Illinois) was responsible for communicating with authors and keeping the entire process running smoothly.

In the editorial review process, Associate Editor Emmy Goldknopf played a key role in reviewing and editing articles. The quality of the encyclopedia owes a great deal to Dr. Goldknopf’s broad knowledge of cognitive science as well as her editorial skills and persistence.

Bringing an encyclopedia of this scope to completion has been a challenge for all of us, but it is hoped that the results will prove useful and illuminating to a wide variety of students of the mind.

Hal Pashler
University of California, San Diego
La Jolla, CA
ACCESS CONSCIOUSNESS

Consciousness, according to the American philosopher Ned Block, is a “mongrel” concept—a conglomeration that picks out a number of very different mental properties that are nevertheless treated as undivided and denoted by a univocal concept. Access consciousness is one of these kinds of consciousness.

This entry will (a) discuss the notion of access consciousness—as distinguished from phenomenal consciousness—in the work of Block; (b) consider the introspective, experimental, and conceptual support for this distinction; (c) briefly consider Block’s contention that much current work on consciousness conflates these two kinds of consciousness; and (d) conclude with a consideration of some criticism that may be directed at Block’s approach.

To see what Block means by access consciousness (A-consciousness), consider the example of driving and being in the cognitive state of perceiving a stationary car on the road ahead of you. This is an A-conscious state inasmuch as its content (that there is a stationary car ahead) is freely available to your cognitive and action-regulating resources—which you may, in this instance, decide to employ in pressing the brake pedal of your own car, planning what to do next, and so on. Characteristically, A-conscious content is linguistically reportable (e.g., “I see that the car ahead of me isn’t moving”), but is also attributable to the lower animals by virtue of their ability to use perceptual content to guide action.

To put it simply, an organism is in an A-conscious state if that state is poised for free use in controlling thought and action for that organism. More formally, A-consciousness consists in the broadcasting of representations for free use in reasoning and for direct rational control of action of the agent (with rational understood in a broad enough sense to include poor reasoning).

The notion of A-consciousness belongs to the family of information processing or functional theories of consciousness—but according to Block’s distinctive account, this is only part of the story. The other main part is phenomenal consciousness (P-consciousness)—the experiential dimension of conscious experience itself, that is, what it is like to experience pain, to see a building, to hear a bell, to smell a flower, and so forth. According to Block, it is not A-consciousness, but P-consciousness that seems to be a scientific mystery. In the vast majority of cases, A-consciousness and P-consciousness are coextensive—such as in the example of seeing the car ahead on the road, where one has both the content available to A-conscious control of thought and action and the P-conscious experience of seeing a large, stationary object ahead. In certain instances, however, a breakup, or disassociation, is thought to appear between the two types of consciousness. We shall consider such cases in more detail later.

In general, Block’s theory of A-consciousness should be viewed as an integral part of a compromise position designed to accommodate both an information processing and a phenomenal view of consciousness. The main thrust of Block’s work has thus been to emphasize the distinction between A-consciousness and P-consciousness and to argue that some current work in psychology, philosophy,
cognitive science, and neuroscience is subject to fallacious inferences deriving from the failure to distinguish all the relevant meanings of consciousness.

**Introspective, Experimental, and Conceptual Support**

The first type of data that Block enlists in support of his account is introspective data, in the sense of “our impressions of how things seem to us.” An example of this is the experience of suddenly noticing a sound (of, say, an electrical appliance or a distant jackhammer) and realizing that this has been going on for a while without one attending to it. Once the sound has been noticed, one is A-conscious of it, but it seems that there was a period before that point where one was in a state of P-consciousness with no additional A-consciousness of the sound. If this is the right way to interpret what happens, it seems that we have a simple disassociation between A-consciousness and P-consciousness.

A second type of data derives from experimental studies and clinical cases, including those of the attentional blink, binocular rivalry, blindsight, the Capgras delusion, Cotard’s syndrome, prosopagnosia, and visual neglect. Prosopagnosia (also known as “face blindness” or facial agnosia) is a neurological disorder characterized by the inability to recognize faces. Yet, under certain experimental conditions, it appears that prosopagnosics have some information about the faces they are seeing, even though this information is A-unconscious. According to Block, it is thus the lack of A-consciousness of the information, rather than the P-conscious lack of the familiarity of the face, that defines the disorder—and attention thus needs to be paid to both dimensions of the condition. Now consider the much-discussed phenomenon of blindsight, a clinical condition associated with brain damage to the visual cortex in which the subject responds to visual stimuli without consciously perceiving them. Specifically, if a stimulus is flashed in a patient’s “blind” area, he will be able to “guess” (with much higher than chance outcome) certain simple features of the stimulus, given highly limited forced choice conditions. Clearly, the patient is unable to use the visual information from the blind area in the access sense of consciousness because he has no ability to deploy the information in reasoning or the rational control of action. Neither, however, does the patient appear to be conscious of the stimulus in the phenomenal sense of consciousness; if we take his word for it, he has no experience of the stimulus. Like the prosopagnosic patient, the blindsight patient thus lacks both the relevant A- and P-consciousness of the stimulus. To get a complete disassociation between the two types of consciousness, Block therefore introduces a conceptual maneuver that will be discussed shortly.

The third type of data employed by Block is entirely imaginary in nature (in a strict sense) and derives from philosophical thought experiments that are assumed to provide conceptually possible cases. One thought experiment is that of superblindness. Here we imagine a blindsight patient with the counterfactual ability to prompt himself at will to guess reliably what is in his blind field. As a result of this remarkable auto-prompting, visual information from the blind field simply pops into his thoughts without any corresponding visual experience of the stimuli. In such a case, the patient would have A-consciousness without the P-conscious visual experience—and we thus have a clear disassociation between the two types of consciousness.

**Conceptual Conflation of A-Consciousness and P-Consciousness**

Relying on the distinction between A-consciousness and P-consciousness, Block has criticized contemporary work on consciousness—including that of Francis Crick and Christof Koch, Daniel Dennett, John Searle, and others—for fallaciously conflating the two notions. One conflation is to analyze P-consciousness in terms of information processing (i.e., A-consciousness); the directly opposite conflation is to analyze A-consciousness in terms of experience (i.e., P-consciousness). Finally, an even more basic confusion occurs if one is operating with an unanalyzed notion of consciousness to which neither A- nor P-consciousness properly applies.

Consider Block’s criticism of Dennett’s hypothesis that human consciousness is largely a product of cultural evolution that is working on our basic biological hardware, a product that becomes imparted to human brains in early developmental training. P-consciousness, Block first points out, cannot be a cultural construction in the way that Dennett suggests. After all, human P-consciousness pertains to the way that things look, sound, and smell to human beings—and this clearly is a basic biological feature of beings like us, not a cultural construction that children have to learn as they grow up. Culture may
have an impact on P-consciousness, but it does not create it. Analogously, culture has an impact on feet (e.g., by making it fashionable to wear constricting shoes), but culture does not create feet. Consider, now, A-consciousness. It, similarly, cannot be a cultural construction because, according to Block, it is a basic biological feature of human beings that we are capable of having states that are poised for free use in controlling thought and action. Culture may influence the kind of content that characteristically becomes A-conscious among a group of people, but it does not create the basic biological ability to make cognitive content available for the control of thought and action. In Block's analysis, then, Dennett's claim about the cultural construction of consciousness is shown to be untenable—and this is because Dennett has confused himself by applying an unanalyzed notion of consciousness rather than the specific concepts of A-consciousness and P-consciousness.

Criticism of the Two Concepts Approach

Block's theory of A-consciousness—and, more broadly, his two-concept distinction—has been subjected to a wide range of criticism from thinkers such as Bernard Baars, David Chalmers, Daniel Dennett, David Rosenthal, and many others.

One line of criticism advanced by Dennett is that Block mistakenly inflates a mere difference in degree between A-consciousness and (what Block calls) P-consciousness into a difference in kind. According to Dennett, Block's distinction can be wholly accommodated in terms of richness of content and degrees of influence—that is, in purely quantitative, A-consciousness terms. Everyday instances of conscious awareness have an immense richness of information, whereas unusual cases, such as the blind field perception of blindsight patients, have a paucity of information. This is simply a difference of quantity, however, not of quality, and it explains the highly delimited but still statistically significant success of blindsight patient "guesses." According to Dennett, if a blindsight patient achieves the superblindside ability to prompt himself at will to guess reliably what is in his blind field, this amounts to the full restoration of that patient's visual consciousness, and that, in turn, impugns Block's strongest case of an A- and P-consciousness split. Rosenthal similarly rejects Block's distinction and proposes to account for the data in terms of higher order thoughts rather than in terms of quantitative differences in degree.

Another line of criticism pertains to the ultimate status of Block's distinction. The problem is that A- and P-consciousness, as Block presents it, never actually come cleanly apart. In prosopagnosia, after all, the subject lacks both the experience of knowing a face and the relevant information about the person. The same is true of blindsight. The blind field information is elicited under forced choice conditions, and the subject has no access to the information in the sense of having it available for free use in controlling thought and action. In neither case do A-consciousness and P-consciousness actually disunite. The jackhammer example is nebulous; attention in some sense is involved here, but it is not at all clear that this phenomenon is best captured by Block's notion of A-consciousness. Indeed, Block grants that there may be no actual cases of A-consciousness without P-consciousness. This still leaves the purely conceptual scenarios with A and P disassociation, but here one needs to be very wary. As with the superblind patient case, one can certainly imagine, say, that it really is the gravitational pull of the moon, not the action of the heart, that drives the circulation of the blood—and this might lead one to make the conceptual distinction between (a) the rhythmic contractions of the heart and (b) the circulatory pumping of the blood. However, this clearly is neither an evidentially warranted nor a clinically fruitful distinction for scientific medicine to adopt.

Finally, as a compromise position, Block's account is open both to general criticism of information processing accounts of consciousness and to general criticism of phenomenal accounts of consciousness.

Christian Beenfeldt

See also Attention and Consciousness; Blindsight; Consciousness and the Unconscious; Mind-Body Problem; Neural Correlates of Consciousness; Perceptual Consciousness and Attention

Further Readings


Block, N. J. (2007). On a confusion about a function of consciousness. *Consciousness, function, and
ACTION AND BODILY MOVEMENT

Human beings are agents: We have the capacity to act, which involves the capacity to cause and to prevent events. Indeed, to perform an action of a particular kind is to cause (or prevent) an event of a particular kind: To kill someone is to cause their death, and to save their life is to prevent it. Actions are attributed to us as agents on the grounds of the events we cause (or prevent); so an action of killing requires that the agent cause a death, an action of breaking a window requires that the agent cause an event of a window’s breaking, and so on.

It may be thought that the capacity to act goes beyond the capacity to cause or prevent events, for acting is sometimes causing or preventing processes or states of affairs or bringing objects into existence; for example, bringing it about that a top is spinning or that a door is closed, or bringing a painting into existence. Even in such cases, however, the initiation of the process, the coming about of the state of affairs, and the object’s coming into existence are events that one causes (or prevents). Therefore, we can broadly characterize our capacity to act as the capacity we have to cause or prevent events. (In what follows, and for ease of exposition, the entry shall talk mainly about causing events, although most of what is said applies to preventing events also.)

This entry explores actions that involve bodily movements. It asks whether the relationship between an action that involves a motion of the body and that motion of the body is one of identity, cause and effect, or something else. It explores the different problems that arise for those various answers if one conceives of actions themselves as events, and suggests that one solution is to abandon the view that actions are events that cause other events in favor of the view that they are causings of events but not themselves events.

Basic Actions and Bodily Movements

Many of the events we cause we cause by moving our bodies. True, we can cause events without moving our bodies at all; for example, one can cause an offense by not acknowledging an acquaintance in public, or one can cause the death of a plant by neglecting to water it. Nonetheless, all the events we cause by acting (as opposed to by failing to act, as in the examples just given), we cause by moving our bodies: We build bridges, drive cars, catch balls, and paint walls by moving our bodies in ways designed to achieve certain outcomes. This does not imply that all we ever do is move our bodies, as, for example, Donald Davidson claims in his paper “Agency,” for we do many more things than that. But it does imply that actions that involve movements of one’s body have a core role in agency.

In general, moving our bodies is something we can do “directly.” That is, although one can move a part of one’s body by moving another, as when I raise my left arm with my right arm, we have the more basic capacity to move our bodies not by doing anything else, that is, to move our bodies directly. For example, I can raise my left arm directly. This capacity to perform what might be called “basic actions” is fundamental to agency.

Now, since moving one’s body is acting, and since acting is causing an event, then it would seem that moving one’s body is causing an event—an event that consists in one’s body’s moving, just as opening a door is causing an event that consists in a door’s opening. Consider the example of turning one’s head. If I turn my head sideways, it seems that I cause an event: the event of my head turning sideways (i.e., the event that is the change from my head facing forward to its facing sideways). This is exactly parallel to my turning a statue sideways: This action is my causing the event of the statue’s turning sideways.

Actions as Events?

Many philosophers have thought that actions, as well as being causings of events, are themselves events, and that the relationship between an action and the event that is caused through that action is one of cause and effect. Thus, they think that, if I
open a door by moving my arm, my action of opening the door is an event that causes the event of the door’s opening; my action is the motion of my arm and its effect is the motion of the door.

This view, however, is problematic when applied to basic actions. For, if my action of raising my arm is an event that has the event of my arm’s rising as an effect, then the question arises: Which event is my (basic) action? Some, for example, Davidson as cited earlier, have argued that, when it comes to basic actions such as the one just mentioned, the event that is the rising of my arm is my action. So he thinks that, although when I open a door there are two events (the door’s opening and a motion of my arm), when I merely raise my arm (directly) there is only one event. However, because raising my arm is causing a rising of my arm, this view has the implausible consequence that sometimes our causing an event and the event thus caused is one and the same, as argued by Maria Alvarez and John Hyman in their 1998 paper “Agents and Their Actions.” Others have suggested that, when I raise my arm, my action is an event that causes, but is distinct from, the event of my arm’s rising. Again, this view generates difficulties: It implies that my actions are either neurophysiological events that happen inside my body and that, for example, cannot be observed without the help of a special instrument or nonphysiological events (perhaps “merely” mental events) that cause motions of my body, such as an arm’s rising. This latter suggestion brings with it difficulties about psychophysical causal interaction, as well as the danger of overdetermination, for it would seem that the bodily motions at issue are certainly also caused by neurophysiological events.

There are other ways of avoiding these difficulties. One is to deny that, when one moves one’s body, one causes one’s body to move. This view is encouraged by the thought that when we move, for example, other objects, we do so by moving our bodies, whereas when we move our bodies, we do it directly. But it is difficult to see why this should imply that moving a book is causing it to move, whereas moving one’s arm is not causing one’s arm to move—rather than that the first is causing the book to move “indirectly” whereas the second is causing one’s arm to move “directly.”

A final option is to abandon the view that actions are themselves events; so my action of moving my arm is not an event that causes the motion of my arm—it is the causing of that motion. On this view, although to act is to cause an event, an action is not itself an event that causes other events but rather the causing of an event by an agent—again as argued by Alvarez and Hyman in their 1998 paper.

Maria Alvarez

See also Explanation of Action; Freedom of Action; Mental Action; Phenomenology of Action; Philosophy of Action

Further Readings


-action slips

The term action slip refers to an action made in error that conflicts with one’s will or conscious intentions. For example, suppose you are arriving at work contemplating the day’s tasks. As you approach your office, you reach for your keys. Only when you are about to insert the key into the lock of your office door do you realize that you have selected your house key and not your office key. As a second example, suppose you normally take two sugars in your coffee. While adding sugar, you become engrossed in conversation with a colleague. A moment later, you realize that you have no idea how many sugars you have added—maybe two, maybe more. Slips such as these seem to occur on a daily basis in the actions of normal adults, particularly when performing habitual or routine tasks such as dressing/grooming, preparing/eating meals, or commuting, and particularly when distracted or interrupted or attention is otherwise occupied. This entry begins by describing behavioral studies and taxonomies of action slips. This is
followed by a review of theoretical and computational accounts of the cognitive processes underlying such slips. The final section discusses the relation between action slips and the errors made by patients with neurological damage.

Behavioral Studies of Action Slips

Action slips were first noted in the psychological literature by William James at the end of the 19th century, but they were not systematically investigated until the pioneering diary studies of James Reason in the late 1970s. Reason asked participants to record details of all of their action slips over periods of a week or two. These studies, together with work shortly after by Don Norman, provided a corpus of action errors that has allowed researchers to categorize action slips into a number of classes or types.

One frequent type of slip is the *capture error*, where actions directed toward an intended goal are “captured” either by a superficially similar sequence of actions with a different goal or by relevant objects in the immediate environment. Reason gives the following example from his study: “I meant to get my car out, but as I passed through the back porch on my way to the garage I stopped to put on my Wellington boots and gardening jacket as if to work in the garden” (Reason, 1979, p. 73). The earlier example concerning the office key might also be seen as an example of a capture error, where the operation of selecting a key is captured by the routine or habit of selecting one's house key.

Further types of error concern the ordering of individual actions within a longer sequence. Thus, an action might be performed before it is appropriate, resulting in an *anticipation error*. Alternatively, an action might be unnecessarily repeated (a *perseverative error*, as in the coffee and sugar example earlier) or left out (an *omission error*). One form of omission error that is particular prevalent is the *post completion error*. This is an error where the omitted step occurs after successful completion of a task's goal. Common examples include leaving the original documents in the photocopier after successfully making the necessary copies, or leaving the goods on the counter of a shop after paying and collecting the change.

Many other types of action slips may be identified, but a final category that is of importance to theoretical accounts of action selection is that of *object substitution*. In such errors, the correct action is performed but in association with the wrong object. The example earlier of the office key can be seen as an example of such an error, but a clearer example provided by Reason is of someone who “put the butter on the draining board and two dirty plates in the fridge, instead of the other way round” (Reason, 1979, p. 72).

The main result of Reason's diary studies was that action slips tend to arise when one is distracted, performing a routine task, or both. Attempts to elaborate on these findings have, however, been beset by methodological difficulties. Thus, although action slips are relatively common in everyday behavior, they are not sufficiently frequent to support direct observational studies—Reason found in his diary studies that participants recorded, on average, less than one action slip per day. Moreover, action slips are difficult to elicit in the laboratory, and although some laboratory studies have attempted to test hypotheses concerning the origins of slips, interpretation of those studies is compromised because it is unclear whether the action errors resulting from experimental manipulations are true action slips. For example, several laboratory studies have sought to elicit action slips by placing participants under attentionally demanding conditions by requiring them to complete two tasks simultaneously or by interrupting them at various points within a task. Such studies have confirmed that errors in action selection are more frequent when participants are attempting two tasks simultaneously than when attempting either task individually, but it is unclear whether the errors are comparable to action slips as seen in everyday action because the experimental tasks necessarily differ from the idiosyncratic everyday routines of the participants. Even if this point is ignored, the studies have also failed to clarify whether different kinds of cognitive failure might be behind the different types of slips.

Theoretical Accounts of the Causes of Action Slips

In everyday discourse, action slips are typically attributed to absent-mindedness. However, William James noted their significance for psychological theories of the control of behavior in his first discussion of errors of action. James distinguished between...
habit and willed or voluntary movement. He noted that without habit we would be reduced to the level of a child performing each action for the first time. Habits thus allow us to direct mental processing to less mundane tasks. But there is a cost to this—actions may at times be inappropriately controlled by habit. At least some action slips (e.g., capture errors) may be argued to be a result of this inappropriate control.

Modern accounts of action slips generally retain the distinction between a system for the control of routine behavior (James’s habits) and one for the control of nonroutine behavior. Action slips then arise when the routine system operates without monitoring or feedback from the nonroutine system, as might occur when the nonroutine system is engaged in some attentionally demanding task. Action slips thus provide insight into the operation of the routine system.

One of the most influential accounts of that routine system is due to Norman who, in collaboration with Tim Shallice, introduced the idea of an action schema—an organized memory unit for the control of a sequence of actions. Within Norman’s activation-trigger-schema (ATS) system, action schemas are held to be hierarchically organized, to have triggering conditions that specify when they might apply, and to have associated activation values that may be affected by several factors, including conscious intention, other schemas, and features of the environment. The hierarchical organization of schemas means that one action schema (e.g., for opening your office door) may have several so-called child schemas (e.g., reaching for keys, selecting a key, placing the key in the lock, and turning the key). Triggering conditions place loose temporal ordering constraints on schemas by specifying the physical situations in which a schema may be selected. Thus, a key cannot be placed in the lock if a key has not already been chosen. Finally, activation values determine which schema or schemas are selected, and hence control behavior, at any point in time.

According to ATS theory, action slips may result either from errors during the formation of an intention or from faulty activation of a schema. Thus, a capture error will arise if three conditions are met: (a) The environment closely resembles that in which the incorrect schema is normally performed (and so the environment activates the schema), (b) the incorrect schema is very well learned (and so its threshold for selection is low), and (c) activation of the intended schema is not sufficiently maintained (e.g., due to distraction). Similarly, omission and anticipation errors arise because of faulty activation of child schemas from the parent schema, leading to a schema’s activation failing to reach threshold or reaching threshold too soon.

**Computational Accounts of Action Slips**

Norman and Shallice developed ATS theory into a fuller account of the functional structure of mind—the contention scheduling/supervisory attentional system theory. Within that theory, routine action is held to be the product of contention scheduling, a system based on ATS theory. In nonroutine situations, the supervisory attentional system may bias contention scheduling by activating appropriate schemas or inhibiting inappropriate ones. Two computational accounts of contention scheduling have since been developed and shown to be capable, under nonideal conditions, of simulating action slips in routine behavior. Both accounts assume that action slips can be explained without a full theory of the supervisory attentional system.

One computational account implements ATS theory by using a hierarchically structured interactive activation network. The network consists of schema control nodes that have associated activation values and that compete for activation through mechanisms standard within the interactive activation framework. Schema control nodes may be excited by top-down influences (from higher level schemas or the supervisory attentional system) and bottom-up influences (the representation of the environment), and inhibited by the control nodes of competing schemas, that is, alternative schemas that achieve the same purpose or that require overlapping cognitive resources. The account extends ATS theory by explicitly incorporating a further interactive activation network whose nodes represent objects on which schemas may act. For example, a schema for choosing a key needs to be related to a set of potential keys. Simulation studies have demonstrated that the interactive activation account of contention scheduling can generate both well-structured error-free routine action and, if activation flow is disturbed through the addition of noise to schema control units or object representation units, action slips of each of the types discussed earlier.
An alternative computational account of the operation of contention scheduling is based on computational framework of simple recurrent networks. Within this framework, representations of inputs (e.g., the objects that are currently visible or held) are combined mathematically according to a learned function with a representation of task context to generate an output action and an updated representation of the task context. By iterating this process, a simple recurrent network model is able to generate a sequence of actions. Like the interactive activation model, the simple recurrent network model is able to generate well-formed error-free action sequences. Moreover, when associations between inputs and outputs are corrupted through the addition of noise to the context representation, the model may drift from one learned sequence to another, thus producing capture errors. If the capturing sequence is the same as the captured sequence, then such errors may appear as omission errors (where a step is skipped) or perseverative errors (where a step is repeated).

Errors in Action Following Neurological Damage

Action errors are common in various groups of neurological patients. Many of these errors resemble the slips and lapses of neurologically healthy individuals, although the patient errors tend to be more frequent and more unusual. Patients have been described with left parietal cortical damage who, when asked to demonstrate how to light a match, hold the wrong end or strike the match against the wrong part of the matchbox. These patients, it has been argued, have an impairment affecting either their action schema knowledge or the links between object representations and the schemas they evoke. Patients with injuries affecting their frontal lobes are also at increased risk of “action disorganization syndrome.” Although there is debate within the literature, the deficit here appears to be one of generating a well-formed goal-directed sequence of actions, with both sequence errors and object substitution errors being relatively frequent. Object substitution errors may also be frequent in the everyday activities of patients with dementing illnesses such as Alzheimer’s disease.

Richard P. Cooper

See also Apraxia; Attention and Action

Further Readings


Addiction

Addiction is a chronic and relapsing disorder marked by persistent substance use despite a host of negative consequences. The U.S. Drug Enforcement Administration’s Controlled Substances Act (CSA) identifies five classes of drugs: narcotics, depressants, stimulants, hallucinogens, and anabolic steroids. Substances within each of these classes, with the possible exception of anabolic steroids, are used to change one’s mood or thoughts by way of altering the delicate chemical balance within the central nervous system. It is thought that all substances of abuse are capable of producing euphoria, and the extent to which this euphoria is reliably produced by the substance increases the possibility of the formation of addiction to it.

The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), defines two primary forms of substance use disorders (SUDs), namely, substance abuse and substance dependence. Substance abuse is marked by (a) recurrent substance use resulting in the failure to fulfill role obligations (e.g., missing work or school due to the consequences of substance use), (b) recurrent use in physically hazardous situations (e.g., drinking and driving), (c) continued use despite negative interpersonal consequences, or (d) repeated substance-related
legal problems. Only one of those symptoms must be present within a 12-month period to warrant the diagnosis of substance abuse. Substance dependence, however, is typically conceptualized as a more severe substance use disorder, which is, in turn, more commonly associated with the term addiction. Substance dependence is marked by the presence of at least three of the following symptoms occurring together within a 12-month period: (a) tolerance (i.e., the need for greater amounts of the substance to achieve the same effect); (b) withdrawal (i.e., a group of symptoms that occur when the substance is abruptly discontinued or taken in reduced amounts); (c) taking the substance in larger amounts or over longer periods of time than intended; (d) persistent desire or unsuccessful effort to stop or cut down substance use; (e) a great deal of time spent in obtaining, using, and recovering from the effects of a substance; (f) giving up important activities because of substance use; and (g) use despite physical or psychological problems caused or exacerbated by the substance. Revisions to the DSM-IV are under way, and the proposed changes include a single disorder category, namely, substance use disorder specified as moderate or severe. In light of the DSM-IV revisions, there continues to be considerable debate in the field as to the best terminology for substance use disorders and whether they should be collectively referred to as addiction, dependence, or an alternative term. Nevertheless, there is clear recognition that addiction represents a chronic and debilitating disorder with enormous costs to the affected individuals, their families, and society.

In this entry, the epidemiology, etiology, and treatment of addiction will be briefly summarized, with important theories and future directions highlighted.

Epidemiology

Substance use disorders, as currently defined in the DSM-IV, represent highly prevalent mental disorders. Based on results from the National Epidemiologic Survey on Alcohol and Related Conditions in 2007, prevalence, or the total number of individuals diagnosed with alcohol abuse within their lifetime and the past 12 months, is estimated at 17.8% and 4.7%, respectively, while the prevalence of lifetime and 12-month alcohol dependence is 12.5% and 3.8%, respectively. Epidemiological data for drug abuse suggest the prevalence of 12-month and lifetime drug abuse to be 1.4% and 7.7%, with drug dependence being less frequent in 12-month and lifetime estimates at 0.6% and 2.6%, respectively. Results consistently demonstrate that SUDs are highly comorbid with one another and with other forms of psychopathology, particularly mood, anxiety, and personality disorders. Recent research has also shown that despite the debilitating nature of SUDs only a small proportion of individuals with drug abuse and less than half of patients with drug dependence ever seek treatment. A similar pattern of infrequent treatment seeking is seen for alcohol dependence, as only a quarter of individuals diagnosed with this disorder report seeking help. Together, these results suggest that substance use disorders (a) are highly prevalent; (b) are debilitating and often comorbid with other forms of psychopathology; and (c) despite their high frequency and associated impairments, treatment seeking for these disorders remains uncommon.

Etiology

The etiology of SUDs is inherently complex and likely involves the interplay between biological and psychosocial factors. At the biological level, two factors that modulate behavior, reinforcement and neuroadaptation, are implicated in the addictive process. Traditionally, it was assumed that addicts are motivated to take drugs either for the pleasure induced by the drug (positive reinforcement model) or to avoid the unpleasant consequences of withdrawal (negative reinforcement model). Drugs of abuse possess acute positive reinforcing effects because of their interactions with specific neurotransmitter systems within the mesolimbic reward circuitry. This circuitry, which encompasses connections between the ventral tegmental area and the basal forebrain (including the nucleus accumbens, olfactory tubercle, frontal cortex, and amygdala), is known to be involved in modulating behavioral responses to stimuli that activate feelings of motivation (reward) and reinforcement through the neurotransmitter dopamine, among other neurotransmitter systems (e.g., opioid peptides). Dopamine’s role in the mesolimbic reward system is widely believed to be crucial for mediating the rewarding effects of drugs, as most drugs of abuse have been found to increase dopamine within the nucleus accumbens.

Recently, however, the positive and negative reinforcement models have been criticized as being
neither necessary nor sufficient for the development of SUDs. The lack of strong withdrawal syndromes present in some common drugs of abuse and the dissociation between subjective levels of pleasure and compulsive drug-taking behavior have been proposed as examples of the limitations inherent in these models of addiction etiology. The incentive-sensitization model of drug addiction, which was developed by Terry Robinson and Kent Berridge in response to these limitations, posits that repeated administration of potential drugs of abuse cause long-lasting neural adaptations in the mesolimbic reward circuitry, which renders the brain reward systems sensitized to drugs and drug-associated stimuli. Thus, although the acute reward and initiation of addiction appear to be dependent on dopamine’s involvement in the mesolimbic reward system, the sensitized brain regions in this model are thought to mediate a subcomponent of reward termed the incentive salience, or craving, rather than the pleasurable effects of drugs per se. Studies have convincingly demonstrated that, with repeated drug exposure, the dopaminergic response becomes triggered by the presentation of drug cues (e.g., the sight of a needle to an individual addicted to heroin or the smell of tobacco smoke to a smoker) as opposed to the administration of the drug itself, which is consistent with the proposed associative learning processes in addiction. In particular, end-stage addiction has been hypothesized to result primarily from the gradual recruitment of the prefrontal cortex and its glutamatergic projections to the nucleus accumbens. This pathological form of neuroplasticity in excitatory transmission is hypothesized to promote the compulsive nature of drug seeking by decreasing the value of natural rewards, reducing cognitive control over drug-related behavior, and enhancing glutamatergic response to drug-associated stimuli.

Genetic factors are widely accepted as determinants of the risk for substance use disorders. Twin and adoption studies have estimated that approximately 50% of the variance in risk for developing alcoholism is due to genetic factors, while the genetic contribution to drug use disorders ranges between 60% and 80%. Studies have also shown that common genetic and environmental factors that influence the vulnerability to substance use disorders are largely, or entirely, nonspecific to any particular drug class. For example, results from the Harvard Twin Study of Substance Abuse revealed that between 50% and 85% of the vulnerability to drug use is common across different categories of illicit drugs. Common genetic and environment risk factors may also underlie the comorbidity between substance use disorders and other psychiatric illnesses such as mood and anxiety disorders. In short, genetic factors account for a substantial portion of the liability to substance use disorders, although these factors are less specific to drug classes and even to psychiatric disorders than originally thought. Although considerable progress has been made in elucidating the relative contribution of genetic and environmental factors to addiction, the identification of specific risk genes remains elusive.

Treatment

Given the complexity of addiction phenomenology, the available treatment approaches are only modestly effective. The standard of care for addiction consists of initial detoxification for cases in which withdrawal symptoms must be medically managed. Following detoxification, intensive inpatient or outpatient programs are typically recommended. A combination of psychotherapy and pharmacotherapy is likely to give patients the best chance of recovery. In terms of psychotherapy, cognitive behavioral approaches have received the most empirical support for their efficacy. A large multisite treatment study called Project MATCH (1997) has also found support for 12-step facilitation and motivation enhancement for the treatment of alcoholism.

Numerous pharmacological treatments have been proposed for the treatment of substance use disorders, yet few have proven effective at maintaining long-term abstinence. Substitution therapies that replace the addictive substance with another, less detrimental substance have been proven to be efficacious for the treatment of opiate addiction; however, these opioid receptor agonists (e.g., methadone) only manage the disorder and may themselves promote the same neuroadaptive processes as the originally abused drug, thereby maintaining the strength of the addiction rather than treating it. The opioid receptor partial agonist, buprenorphine, has shown greater efficacy for the treatment of opiate dependence and can be administered on an outpatient basis. The Food and Drug Administration (FDA) has approved three medications for the indication of alcoholism, namely, naltrexone, acamprosate, and disulfiram. The opioid
antagonist naltrexone and the partial glutamate co-
agonist acamprosate target the state of craving by
blocking the opioid receptors and NMDA glutamate
receptors within the reward circuit, respectively.
Additionally, a variety of antidepressants have been
used in the treatment of addiction, based in part on the
assumption that they will alleviate some of the anhe-
donia, or inability to experience pleasure, reported by
patients during protracted withdrawal. Bupropion
is one such antidepressant that has been reported
to increase abstinence from cigarettes in smokers.
More recently, immunotherapies (i.e., vaccines) are
being developed for the treatment of cocaine abuse,
which stimulate the body’s own immune system to
attack the substance, thereby reducing the amount
of cocaine that reaches the brain. Although promising,
these immunotherapies fail to reverse the neuro-
adaptive processes responsible for addiction, and thus
craving for the drug may remain despite treatment.

Summary
Addiction is a chronic and relapsing condition
caused by the interplay between genetic and envi-
ronmental factors. Research on the neurobiologi-
cal bases of addiction has consistently highlighted
the ability of drugs of abuse to potently activate the
mesolimbic dopamine system, producing an intense
subjective feeling of reward. Although various drugs
of abuse have different pharmacological properties
and associated intoxication effects, their common
activation of this reward circuitry subsequently pro-
duces the compulsive drug seeking that characterizes
this disorder. Addictive disorders are highly frequent
in modern society and associated with devastating
psychosocial and health consequences. These dis-
orders often co-occur with other psychiatric illness,
such as mood, anxiety, and personality disorders.
The currently available treatments for addiction are
only moderately effective, and the combination of
pharmacotherapy with behavioral therapy is often
recommended. Nevertheless, the pharmacological
treatment options for various substance use disor-
ders are scarce relative to other psychological dis-
orders, and there is a strong need to develop more
effective medications as well as to identify the bio-
logical and psychological factors within patients
that can guide the selection of effective treatment
approaches at the individual level.

Lara A. Ray and Kelly E. Courtney

See also Behavioral Therapy; Reinforcement Learning,
Psychological Perspectives

Further Readings
Compton, W. M., Thomas, Y. F., Stinson, F. S., & Grant, B.
F. (2007). Prevalence, correlates, disability, and
comorbidity of DSM-IV drug abuse and dependence in
the United States: Results from the National
Epidemiologic Survey on Alcohol and Related
Hasin, D. S., Stinson, F. S., Ogburn, E., & Grant, B. F.
(2007). Prevalence, correlates, disability, and comorbidity
of DSM-IV alcohol abuse and dependence in the United
States: Results from the National Epidemiologic Survey
on Alcohol and Related Conditions. Archives of General
Psychiatry, 64, 830–842.
of addiction: A pathology of motivation and choice.
Koob, G. F., & Le Moal, M. (2001). Drug addiction,
dysregulation of reward, and allostatic
Neuropsychopharmacology, 24,97–129.
treatment with adult alcohol and illicit drug users: A
meta-analysis of randomized controlled trials. Journal of
Studies on Alcohol and Drugs, 70, 516–527.
word? Addiction versus dependence in DSM-V.
and neurobiology of addiction: An incentive–
Tsuang, M. T., Bar, J. L., Harley, R. M., & Lyons, M. J.
(2001). The Harvard Twin Study of Substance Abuse:
What we have learned. Harvard Review of Psychiatry,
9, 267–279.

Affective Forecasting
When deciding whether to order a pizza or a cheese-
burger, to live in California or Winnipeg, or to make
the big leap to have children, people may spontane-
ously envision the emotions that each possible future
would evoke. The capacity to make such affective
forecasts stems from humans’ exceptional ability to
step back from their present experience and men-
tally travel through time to re-live or “pre-live” per-
sonal events. However, our emotional projections
can often be biased. This entry introduces the notion
of forecasting errors and the two main reasons behind them.

Affective forecasts have been shown to predict a variety of decisions, from whether to comply with a persuasion attempt to whether to get a flu shot. These decisions may be flawed, however, given that a large body of evidence shows that individuals often stumble in predicting their affective reactions to future events. People frequently overestimate how happy they will be after positive events, such as getting promoted, and how sad they will feel after negative events, such as the loss of a favorite sports team or political candidate. This tendency to overestimate the intensity and duration of emotional responses (termed the impact bias) is perhaps the most commonly observed forecasting error. Yet, there are also important cases in which people underestimate their emotional responses to events. For example, individuals in committed romantic relationships have been shown to underestimate the pleasure they would derive from interacting with a stranger of the opposite sex. Finally, in some cases, people may even mispredict whether the emotional consequences of an event will be positive or negative; as one example, most people expect that they would feel better if they had the opportunity to punish a free rider, even when exacting revenge accentuates distress.

Although forecasting errors can assume a wide variety of forms, most of these errors stem from the failure to fully appreciate two fundamental principles of emotion. First, emotions are transient; although a given event may provoke a strong emotional response at first, even powerful emotions are likely to fade away quickly under most circumstances. Some of the most common forecasting errors spring from the failure to appreciate the power and dynamics of such hedonic adaptation. Second, context matters; both the emotions we experience and those we imagine depend on the context in which the experiencing and the imagining occur. Overlooking the importance of context can unleash a host of forecasting errors—particularly when emotions are imagined and experienced in very different contexts.

**Hedonic Adaptation**

An important early discovery in the study of affective forecasting was that people show a striking blind spot for their capacity to adapt to negative events. Specifically, individuals often fail to appreciate the effectiveness of their “psychological immune system,” that is, the array of defense mechanisms and emotion regulation strategies that help them adapt quickly and effectively to upsetting events (an oversight labeled immune neglect). People also tend to overlook the extent to which they will make sense of extraordinary events, such that the events come to seem ordinary and unremarkable. This sense-making process can deprive both positive and negative events of their emotional power—to a degree that most people seem unable to anticipate ahead of time.

Whereas overlooking the existence of hedonic adaptation consistently leads people to make excessively extreme affective forecasts, overlooking the factors that speed up or short-circuit adaptation can beget a variety of other forecasting errors. For example, punishing a free rider may prevent people from “moving on,” thereby interrupting the adaptation process, such that negative feelings are actually more powerful than expected. Adaptation may also be short-circuited by uncertainty; when people are uncertain why a positive event occurred, they have trouble making sense of it and tend to experience more lasting pleasure. Yet, because people fail to appreciate the dynamics of adaptation, they may seek to eliminate uncertainty, thereby eliminating an important source of lasting pleasure.

**Context**

A second fundamental source of forecasting error lies in the failure to appreciate the power of context. When imagining a future situation, people tend to pay too much attention to the central event, such as the victory of their favorite football team, while overlooking other situational factors that will moderate its impact, such as traffic jams, mosquitoes, and phone calls from chatty relatives (a phenomenon termed focalism). The broad tendency to neglect the power of context becomes particularly problematic when emotions are imagined and experienced in very different psychological contexts. People are often in a different state of physiological arousal when they make their forecast than when they experience the event, laying the ground for projection biases (or empathy gaps). That is, individuals in a “cold state” (e.g., unaroused, sated, or bored) have considerable difficulty predicting how they will feel when they are in a “hot state” (e.g., aroused, hungry, or curious).

Even in the absence of such hot state/cold state differences, people often find themselves making
choices—and affective forecasts—in a context that differs from the context in which they will actually experience the outcome of their choice; affective forecasts are commonly made in a “joint-evaluation” mode, in which people compare multiple options in parallel, whereas the outcome of the decision is usually appraised in a “single-evaluation” mode, in which only the chosen option is available (creating a distinction bias). When choosing a television, for example, one is faced with a wide array of options, potentially magnifying the minor differences between them, but those differences are unlikely to remain salient while enjoying the selected television in one’s home, where the chosen TV is removed from its former competitors. This difference in evaluation modes is particularly problematic as people tend to focus on the differences between options and ignore their common features when forecasting, even when these features are key elements for enjoyment when experiencing. Thus, the broad tendency to overlook the power of context—and the ways in which the context will differ when emotions are imagined versus experienced—provides the wellspring for a host of forecasting errors. Even if we are looking at the stars, then, the fact that we are lying in the gutter may impair our mental time travel machines from taking flight.

Jordi Quoidbach and Elizabeth W. Dunn

See also Emotion Regulation; Self-Consciousness

Further Readings


AGING, MEMORY, AND INFORMATION PROCESSING SPEED

This entry describes how the speed with which people can make decisions and their ability to remember events change with age. These abilities are discussed together because behavioral measures of speed and memory efficiency and the rates at which they change with age are very strongly correlated. This once promoted the hypothesis that age changes in all or most abilities are driven by “general slowing” of information processing, possibly due to changes in efficiency of neurons and their ability to interact with each other. The entry then considers how information received by the sense organs is, in turn, recorded and transformed in successive stages of the memory system (i.e., immediate memory, working memory, and long-term memory). The entry also discusses the extent to which each stage depends on information processing. Long-term memory for past life events is addressed, as well as how intentions and plans are held in “prospective memory,” to be evoked and implemented when necessary. Finally, the entry discusses how new information from brain imaging allows us to dissociate the effects of gross brain changes on decision speed, on memory, and on the formation and execution of intentions and plans, suggesting a new interpretation for the apparent strong dependency of memory efficiency on information processing speed.

Measures Sensitive to Aging

The behavioral measurements that are most sensitive to changes in aging brains are memory failures and slowing of the speed with which we can make simple decisions. Slowing seems to reflect a basic change in brain efficiency rather than loss of motivation or greater caution, because the limiting speed at which accurate decisions can be made becomes longer as brain aging progresses.

Until the 1980s, when brain imaging began to provide direct measures of neurophysiological changes in living and functioning brains, speculative models for cognitive aging could be derived only by comparing different groups of older and younger people or by repeatedly testing the same people, as they aged, on the same behavioral tasks. Unfortunately, any behavioral task can give us only two different measures of performance: how fast and how accurately we can perform it. These two measures trade off against each other, as in memory tasks, where, although the immediate measure of performance is the number of errors made, shortening the time that we are allowed to inspect items impairs our ability to remember them. Fergus Craik’s excellent metaphor is that we increase the reliability of our memory in proportion to the “depth of processing” that we can achieve by making connections.
with other material that we already know. Deeper processing takes longer and, because they process information more slowly, older people are at a proportional disadvantage.

For this reason, slowing of information processing speed must, inevitably, appear a sensitive “master” index of age-related change in all tasks. Because people can maintain accuracy by responding more slowly, changes in speed are usually detectable before changes in accuracy are observed. Thus, to be detectable, any observed changes in task performance with age must involve changes in speed. The deeper question is whether, apart from being the critical one of the only two task performance indices that we can measure, speed is also the predominant performance characteristic of our cognitive systems. That is, whether slowing of decision speed is the functional cause of changes in all or most other mental abilities. If it is, we may hope that, by comparing the relative sizes of differences in speed between older and younger people across a variety of tasks, we may be able to use behavioral evidence alone to frame useful hypotheses as to whether aging affects some functional systems earlier and more than others. For example, we might suppose, with the earliest investigators of human behavior such as Wilhelm Wundt and Franciscus Donders, that loss of speed may directly reflect changes in the basic neurophysiological performance characteristics of our brains and central nervous systems. Until the late 20th century, investigators such as Hans Eysenck, Art Jensen, Tim Salthouse, and others suggested that speed on simple tasks directly mirrors elementary biological properties of our nervous systems such as synaptic conduction times, rates of neural transmission, or the degree of degradation of connectivity in neural networks; discussions in cognitive gerontology over the past 50 years have been dominated by this point of view. More recent evidence suggests that declines in maximum information processing speed limit memory efficiency not only by restricting the depth to which we can process information or affecting the maximum rate of rehearsal but also because slowing is a marker for basic neuronal changes in the entire brain that affect all cognitive abilities.

**Memory Stages and Aging**

The frequency and alarm with which elderly people insist that their memories are becoming increasingly unreliable are second only to their complaints of arthritis. This worry is partly misplaced. People tend to overestimate the changes that they experience because they often lack objective external standards of comparison for their own mental abilities. Their subjective impressions tend to overestimate the true extent of the changes that actually occur.

To guide their interpretations of data from behavioral tasks, psychologists have adopted a working assumption that types of activity that must logically occur in temporal sequence are probably supported by separate functional systems. So, tracking the passage of information from our sense organs to our long-term memories, we can consider the first stage of iconic memory, a term for the retention of complete images received by the eye that remain available for 50 milliseconds or so until obliterated by new visual events. After this stage, which seems not to be impaired by increasing age, we process information from these iconic images to extract relevant from irrelevant details. Our resulting perceptions must then be briefly held for assessment against previously acquired expectations or intentions, or against further new information, and possibly internally played back (or “rehearsed”) and then stored for periods varying from seconds to decades and finally retrieved and reinterpreted. Working within the assumption of functional differences between temporally successive processes, it has been natural to design behavioral experiments to test whether old age affects some of these hypothetically entrained processes of input, brief retention, longer term storage, and retrieval more than others.

To find the limits of immediate storage of perceived information, investigators first measured the number of items, such as decimal digits, letters, or words, presented at measured rates and recalled by study participants in the exact order in which they were heard or read. George Miller first pointed out that such “immediate memory spans” for different kinds of material (e.g., digits, words, or letters) seemed to vary within a narrow range of 7 ± 2 (seven plus or minus two) depending on item information load of items computed as the binary logarithm of the number of possible alternatives from which they were drawn. So, because there are fewer possible digits and letters of the alphabet than words, spans for lists of random digits or letters are proportionately longer than spans for random words. The ratios of possible alternative items to those actually presented is not the whole story because, as Alan Baddeley
Aging, Memory, and Information Processing Speed

and Graham Hitch showed, the numbers of syllables that words contain set limits on our capacity: We can immediately “play back” fewer long than short words. The particular physical properties of the sounds we try to remember are also important. Reuben Conrad showed that similar sounding letters such as S and F or A and K are confused in sequences, so that spans for acoustically confusable items are shorter than for different sounding items. Immediate memory spans seem almost unaffected by healthy old age, but they are reduced by pathologies such as dementias. They can be extended by familiarity with material, such as telephone dialing codes or license plates; individuals who master ingenious mnemonic coding strategies are able to achieve indefinitely long “spans.” Thus, actual age changes in the limits of immediate memory capacity can be masked by practice with particular kinds of material and by use of mnemonic strategies. However, age also markedly reduces the speed and efficiency with which mnemonic strategies can be implemented.

At all ages, memory spans are reduced if the items retained are not simply given back in the order in which they were presented but must be reorganized or transformed, as when reported in reverse order (backward memory span). The number of items that we can reorganize or otherwise transform before recalling them declines sharply as we grow older. The insight that items held in immediate memory are typically not just passively stored for subsequent readout but, rather, dynamically and selectively rehearsed and restructured was captured by Alan Baddeley as the concept of working memory. Baddeley showed that these dynamic processes make demands on information processing speed and on the ability to relate current to past information. He also found that limits to working memory are determined by a trade-off between the rate at which traces of the immediate physical characteristics of sensory input decay and the speed with which they can be “read off” and converted into other forms of mnemonic representations. There is also evidence that functionally separate systems of transient, dynamic storage are used for acoustic representations of words or syllables (the articulatory loop) and for visual information (the scratch pad). It is confidently known that age-related slowing impairs read-off and rehearsal of input, but not yet whether it affects one hypothetical sense-organ-specific system more than another.

To actively select, reorganize, or interpret information briefly held in working memory, we must be guided by some plan or intention as to what to do with it. The formulation, comparison, adoption, and maintenance of goals and plans to guide interactions with the world are termed central executive processes and are known to depend on integrity of the prefrontal cortex in which gross age-related changes indexed by cell loss and white matter lesions appear earlier and are greater than in other cortical structures. Growing evidence shows that normal brain aging affects these higher order processes of attention, intention, and planning in specific ways that are independent of its effects on information processing speed.

The amount of information that can be held in dynamic working memory is reduced by slowing of information processing. Also, because working memory has limited capacity, it must be continually updated by new events in our rapidly changing environments. The outcomes of earlier analyses must be transferred to some other system. Early work assumed that storage is passive rather than active, but later studies showed that remembered information is continuously and radically integrated with, and thus altered by, previously acquired knowledge of the world and also by new information. It is still unclear whether information stored in long-term memory remains unaltered until it is retrieved for further use (recalled) or whether the long-term storage systems are not just passive repositories but continuously self-reorganize their contents, even without conscious sifting and updating. The second hypothesis is currently preferred. For interpretations of age-related changes, two connected issues are the rates of progressive loss of once accurately stored memories and the consequently increasing failure to distinguish between inferred and remembered information. As far as we know, the limits to long-term memory capacity are so large that we do not run the risk of running out of “memory storage space” during our lifetimes. However, it is clear that, for people of all ages, the longer information is stored without being refreshed by reuse, the less accurately it can be retrieved. As years pass and we grow older, we become increasingly aware of gaps in knowledge that was once perfectly recalled. Knowledge intensively studied in youth, such as high school Spanish or university psychology, becomes increasingly irretrievable over periods of a decade or longer, perhaps
The honesty of individuals’ subjective impressions of vividness, reliability, and detail is not in question, but the idea that these recollections, picturesquely termed flashbulb memories, are exceptionally veridical, detailed, and durable is mistaken. When respondents are asked repeatedly to recall striking moments over long intervals, losses and distortions become increasingly evident at any age and seem to become more frequent as old age progresses.

Age impairs the ability to recall information: that is, to have a conscious and reliable conviction of accurate detailed recall of an experience or item of information and, most particularly, to be able to correctly provide some context in which it was experienced. In contrast, recognition of previously encountered information is necessarily easier because it does not demand accurate reconstruction of all of the information originally presented or of the context in which it was encountered. Because recognition requires less information to be stored than does complete detailed recall, it is understandable that it should be less affected by aging. However, differences in accuracy of recall and recognition also seem to reflect deeper differences in underlying function. Old age reduces the probability of confident conscious recall both of items of information and, more particularly, of the contexts in which they were first encountered (explicit memory). But, as Larry Jacoby and Alan Parkin have shown, age seems to have little or no effect on the ability to recognize an item as “familiar,” without conscious and explicit knowledge of specifically when and where it was last encountered (implicit memory).

Loss of accuracy in recalling incidental information about events or statements, such as where they occurred or who uttered them, is a frequent inconvenience in everyday life. This becomes increasingly troublesome in old age. Elderly persons are less able to retain source memory, recalling whether words were spoken by one of two experimenters, or whether they had themselves thought of them; whether they had actually performed, or had only imagined performing, actions or had watched others performing them. Experimenters asking older and younger adults to learn lists of names of actual and fictitious celebrities found that, when these were recalled, the old more often than the young confused “real” with “imaginary” persons. Impairment of the ability to retain the extra information needed to establish the context and source in which information is encountered evidently reduces the reliability of eyewitness testimony of events.
A key theme of research has been whether age changes in memory are entirely or only partly associated with slowing of information processing speed that affects efficiency of input, encoding, and retrieval. A different issue is the question of whether or not memory traces of different kinds of information, such as images, spatial relationships, words, actions, and events, have different and separate representations in the brain and are also differently affected by increasing age. Allowing for the difficulties in computing differences in the information load and complexity of different kinds of material, it seems that, in old age, memory for visual images is less impaired than is memory for words. Perhaps this is partly because recall of visual information has typically been assessed by recognition, because older people have particular problems in drawing or otherwise reproducing or describing pictures. Differences in accuracy of recall of other, different kinds of information have not yet been thoroughly studied.

For humans, as for other animals, the most important evolutionary benefit gained from a capacious and exact memory is not the Proustian luxury of detailed and sensual recollection of the past but the ability to use previous experiences to predict the immediate future and provide appropriate plans and procedures to deal with it. The evidence on age changes in the use of prospective memory to plan and guide future actions is still incomplete. Older adults seem to fail no more often than the young in situations in which there is no direct cue to implement an instruction, such as when they are asked to press a key every 10 minutes, or to regularly turn away from a different, engrossing task to check a clock. The old are also no worse in obeying an instruction to respond to predesignated targets as soon as they appear. In sum, the more an activity is supported by cues available in the immediate environment, the smaller are age differences. Elizabeth Maylor found that when appropriate environmental cues were not available, her older respondents tended to cope by deliberately structuring their lives, even arbitrarily, by rearranging furniture, to remind them of actions that she asked them to remember to carry out.

Evidence From Reaction Times

We have noted that limitations of behavioral measurement imply that measurements of age changes in memory are inevitably confounded with losses in information processing speed. Of all behavioral indices, the speed with which we can make very simple decisions (reaction time [RT] tasks) have probably been the most intensively studied. Consequently, nearly all experimental methodologies and theoretical questions used to study RTs in young adults have been extended for age comparisons. In general, older people are slower at decisions of all kinds and are more affected by any factors that make tasks harder. So, in absolute terms, older participants are more affected by increases in task information load, as assessed from the number of alternative signals and responses among which they have to choose; by reduced signal brightness or loudness; by greater complexity in mapping signals on to responses; and by the size and complexity of sets of signals that have to be discriminated from each other. In all of these situations, the more demanding the task, the greater, in absolute terms, is the difference between the average RTs of groups of old and young adults. In absolute terms, older people benefit more than young adults when successive signals require the same responses (the signal and response repetition effect). They also need longer to prepare to identify signals and program responses to them, and are less efficient at recognizing and using the statistical structure of the distributions of relative probabilities of the moments of appearance of signals to maximize their readiness to cope with them.

The general findings that age slows harder decisions more than easier decisions led to hope that this information could tell us why some cognitive processes are more age sensitive than others. This was undermined by a finding by Joseph Brinley, confirmed by John Cerella. Across a wide range of very different experimental tasks, average RTs for groups of older individuals can be accurately derived by multiplying average RTs for groups of young adults by a simple constant in the range 1.2 to 1.7 (depending on the average age difference between the groups compared). Later studies found that on nearly all tasks so far examined, distributions of RTs for groups of older people can also be precisely matched by multiplying distributions for groups of younger people by the same simple constants between 1.2 and 1.7 (depending on differences in their average ages and conditions). This is vexing because, if some kinds of decisions were slowed more than others, we could argue that the particular processes they involve are particularly vulnerable to aging. Consequently, the generality and theoretical interpretation of this striking regularity have been vigorously questioned.
However, its pervasiveness has led to the suggestion that age-related changes in the brain lead to “general slowing” that affects all cognitive functions in the same proportion; thus, it may be regarded as the prime functional driver, rather than just a symptom or “biological marker,” for all age-related changes in mental abilities. Evidence from brain imaging, discussed in the next section, now shows this to have been a premature judgment.

RT tasks make a different and possibly more interesting point about age differences. The old are not just slower on average but also more variable from trial to trial. It follows that, because RTs obtained from the same person on different occasions can be regarded as samples drawn from a distribution of all of the RTs that the person has made, any increases in the standard error of RT distributions obtained on any single session must mirror a corresponding increase in the standard errors of the means of these different samples of RTs obtained on different occasions. In other words, linear age-scaling of RTs also means increases in variability of RTs from moment to moment and also in variability of mean RTs obtained at different times. Further, because each older person’s mean RTs vary more from session to session, older individuals will necessarily appear to vary more with respect to each other than do the young when compared on any particular occasion. This change in performance cannot be interpreted as another example of “general slowing” and must reflect some other functional change that occurs in aging brains.

A different reason why slowing of information processing has been a central topic in cognitive gerontology over the past 30 years derives from discoveries that variance between older and younger individuals’ RTs on simple laboratory tasks typically accounts for all, or most, of the variance between them on other cognitive tasks, especially simple memory tasks. This motivated Tim Salthouse and Jim Birren’s differently formulated suggestions that “general slowing” is not just a sensitive marker of changes in all other mental abilities but is actually the principal driver of declines in intelligence and memory.

Cognitive gerontologists hope not only to describe changes in abilities that occur in old age but also to understand how aging of the brain causes them. Because all laboratory tasks, including those designed to diagnose local brain damage are, directly or indirectly, also measures of information processing speed, behavioral data alone do not yet confirm this hypothesis. For example, scores on widely used behavioral diagnostic tests for frontal damage such the Stroop or Trails tests are, actually, RTs to unusual stimuli, and the proportions by which these RTs are slowed by age differ only barely detectably (Paul Verhaegen) or not at all (Tim Salthouse, Patrick Rabbitt) from those observed in other, simpler tasks. To the extent that the relative (rather than absolute) extent of age differences in speed of performance are unaffected by the diverse demands of diagnostic tests of local brain damage, these tests cannot distinguish local from global age changes in brain function. For this, we need to examine relationships between age-related changes in behavior and in neurophysiology rather than simply between calendar age and behavior. For example, without direct neurophysiological information from methodologies such as brain imaging, behavioral comparisons do not provide unambiguous evidence for the reasonable assumption that, because age affects the frontal lobes earlier than other parts of the cortex, frontal abilities such as the setting of goals and the development and prosecution of plans to achieve them, suffer correspondingly early impairments. A possible exception to the ambiguity of the behavioral evidence for “frontal aging” is the finding that the elderly are slowed disproportionately more than young adults when abruptly switching between different subtasks or when attempting to do two things at once. The fact that task switching and multitasking depend on the integrity of the frontal lobes supports theories of frontal aging. Further support for the frontal aging theory is Lynn Hasher and Rose Zack’s evidence that tasks demanding continuous control of information held and transformed in working memory show more marked age-related decrements than can be explained by global slowing alone.

**Evidence From Brain Imaging**

Improvements in brain imaging now allow us to directly relate the amounts of observed behavioral changes to individual differences in patterns of local brain changes. We can also test whether brain changes underlying changes in behavioral performance are local as well as general and diffuse.

Brain scanning has also clarified that slowing of information processing drives age-related changes
in different cognitive functions to different extents. Gross brain changes, especially the increase in incidence of white matter lesions and global losses in cortical volume, affect information processing speed much more than other abilities, such as performance on intelligence tests. However, statistical methods such as structural equation modeling show that, although the effects of these gross brain changes on memory are indeed mainly accounted for by concomitant slowing of information processing speed, this is much less true of scores on frontal tests and not at all true of scores on tests of fluid intelligence. This provokes a new interpretation of the very strong relationships between scores on tests of information processing speed and on tests of all other cognitive abilities. Scores on tests of speed are clearly exceptionally sensitive indices of the extent, and progress, of gross changes in the entire brain. In this sense, they are also, inevitably, sensitive markers for the amounts and progress of all of the changes in the brain that bring about losses of all other mental abilities. This proposition that slowing of information processing speed is a marker for brain changes that affect all tasks is clearly logically distinct from the Birren and Salthouse speculation that changes in information processing speed are functional drivers for corresponding changes in performance of all other tasks. This allows reconciliation rather than eliminative confrontation between views. Both perspectives are useful. Specific functional dependencies between information processing speed and, for example, working memory efficiency or improvement due to depth of processing are unchallenged. But we must also bear in mind that information processing speed is a strong symptom and thus an accurate marker for gross brain efficiency. The relative amounts of variance in performance due to direct and merely symptomatic status of information processing speed is a new and basic question that remains to be investigated.

The increasing availability of brain imaging is helpful for investigators attempting to resolve another key methodological problem in studying how age affects cognition. It has long been recognized that, because individuals age at very different rates, calendar age is a relatively weak predictor of cognitive abilities. This raises obvious problems for attempts to study age-related changes by comparing small groups of older and younger people because, within any older sample, the particular brain changes whose effects are under investigation may be present in increasing numbers of individuals but not in all or even in most. Individual differences in patterns of age-related brain changes observed in any group are now known to be considerable and to become more marked as the group ages. Because of the variability of brain changes, variability between individuals’ behavioral data are certain to be inconclusive when, as has been the norm over the past 75 years, comparisons involve small groups of only 15 to 50 elderly individuals. Scanning data now allows behavioral changes to be related to different amounts and patterns of age-related brain changes rather than to the differences in calendar age, which, as we now know, is an imprecise guide to the extent and nature of individual changes in cognitive function.

Patrick Michael Rabbitt

See also Intelligence, Neural Basis; Intelligence and Working Memory; Memory, Neural Basis; Reaction Time; Working Memory

Further Readings

Allais Paradox

The Allais paradox is a paradox in risky choice first proposed by Maurice Allais in the 1950s to challenge the then-dominant view that humans are rational economic actors. According to expected utility theory, a widely used economic theory of choice, humans make decisions between options with uncertain outcomes by considering the probability and value of each possible outcome. Decision makers who choose according to expected utility theory are economically rational: They will display consistent
preferences, and their choices will maximize their own self-interest.

The Allais paradox was one of the first demonstrations that human decision making systematically violates expected utility theory, suggesting that standard models of economic rationality may not be sufficient to explain actual decision behavior. Multiple theories have been put forward to account for the Allais paradox, including theories about the way probability is used in decisions and theories about the role of emotions in decisions. As one of the earliest demonstrations of the limits of the rational actor hypothesis, the Allais paradox helped motivate the psychological study of decision making.

**Expected Utility Theory**

In the 18th century, Daniel Bernoulli proposed the expected utility theory. Bernoulli suggested evaluating a risky prospect by multiplying the probability of each possible outcome by the *utility* of the outcome (a numerical estimate of the value of the outcome to the decision maker). Summing these values over all possible outcomes produces the *expected utility* of an option, and the option with the highest expected utility is chosen. In the first half of the 20th century, Frank Ramsey, John von Neumann and Oskar Morgenstern, and Leonard J. Savage demonstrated that expected utility theory can be derived from a small set of intuitive axioms and that it is the decision strategy that maximizes utility when used over time.

**The Independence Axiom**

One of the axioms from which expected utility theory is derived is the *independence axiom*, which says that any outcome common to two options is irrelevant when choosing between them. For example, consider the following gambles:

<table>
<thead>
<tr>
<th>Pair One</th>
<th>Ticket A</th>
<th>Ticket B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heads: a trip to Rome</td>
<td>10% probability of $5 million</td>
<td>11% probability of $1 million</td>
</tr>
<tr>
<td>Tails: a trip to New York</td>
<td>90% probability of $0</td>
<td>89% probability of $0</td>
</tr>
</tbody>
</table>

Allais proposed, and subsequent research confirms, that decision makers will generally choose Ticket B, $1 million for sure.

Now consider the following gambles:

<table>
<thead>
<tr>
<th>Pair Two</th>
<th>Ticket A</th>
<th>Ticket B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heads: a trip to Rome</td>
<td>10% probability of $5 million</td>
<td>11% probability of $1 million</td>
</tr>
<tr>
<td>Tails: a trip to New York</td>
<td>90% probability of $0</td>
<td>89% probability of $0</td>
</tr>
</tbody>
</table>

In this pair of gambles, decision makers generally prefer Ticket A, a 10% chance of $5 million.

Expected utility theory says that decision makers should choose the same ticket in both pairs of gambles. Pair Two is obtained from Pair One by changing an 89% chance of $1 million to an 89% chance of $0 in both gambles. By reversing their preferences between the two pairs, decision makers are violating the independence axiom. To see why, suppose the gambles are resolved by drawing balls from an urn:
Just as New York was irrelevant to the choice between the vacation gambles, the outcome of drawing balls 12 through 100 is irrelevant to the choice between the Allais paradox gambles. Within each pair, decision makers who draw balls 12 through 100 will receive the same outcome regardless of whether they chose Ticket A or Ticket B. Thus, only the outcomes of drawing balls 1 through 11 are relevant to the decision. Because these are the same in both pairs of choices, decision makers should choose the same ticket in both pairs.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>The Allais paradox resolved by drawing balls from an urn</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Pair One</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Ticket A</strong></td>
</tr>
<tr>
<td>Balls 1–10</td>
<td>$5 million</td>
</tr>
<tr>
<td>Ball 11</td>
<td>$0</td>
</tr>
<tr>
<td>Balls 12–100</td>
<td>$1 million</td>
</tr>
</tbody>
</table>

Explanations for the Allais Paradox

The Allais paradox is generally considered a form of the certainty effect, in which decision makers overweight certain outcomes (i.e., outcomes that are certain) when making decisions. In Pair One, the decision maker is drawn to the certain payout offered by Ticket B. In Pair Two, neither ticket offers a certain outcome, so the decision maker chooses the higher payout Ticket A.

Prospect theory is the most common theory put forward to explain the Allais paradox and other violations of expected utility theory. Prospect theory explains the Allais paradox, and the certainty effect more generally, via a probability weighting function. Because this probability weighting function is steeper near the ends than in the middle, decision makers treat the difference between 0% and 1% as if it were much larger than the difference between 89% and 90%. Thus, a 1% difference in the probability of winning nothing has a much bigger impact in Pair One than it does in Pair Two.

Another common explanation for the Allais paradox is that it is driven by emotion. When choosing between the gambles in Pair One, the decision makers may anticipate the regret they would feel if they took a chance on the risky Ticket A and lost. Decision affect theories propose that decision makers choose certain outcomes to avoid the risk of regretting their decision. In Pair One, decision makers who choose Ticket A and lose know they would have won had they chosen the certain Ticket B, but those who choose Ticket B will not know what the result of choosing risky Ticket A would have been. However, in Pair Two, decision makers who choose Ticket A and lose know they would probably still have lost had they chosen Ticket B. Because they do not anticipate regretting choosing Ticket A in Pair Two, they are more likely to choose Ticket A in Pair Two than in Pair One.

Bethany Weber

See also Decision Making, Neural Underpinnings; Decision Theory, Philosophical Perspectives; Neuroeconomics

Further Readings


Amnesia is a memory disorder characterized by a profound impairment in establishing new memories of people, places, and events and a deficit in retrieving information acquired prior to the onset of illness. It is important to note that amnesia is not an all-or-none condition—it is functionally diverse and its severity varies based on etiology and lesion extent. However, despite the heterogeneity observed across patients, some fundamental characteristics can be established. This entry presents an overview of the syndrome, including the different subtypes of amnesias and their behavioral and neurological characteristics.

Broadly speaking, amnesic patients are unable to consciously bring to mind information encountered after the onset of injury (anterograde amnesia), despite preserved general intelligence and other cognitive functions, such as language. In this sense, their cognitive deficit is selective (i.e., limited to memory). The memory impairment is often described as “global” because patients’ inability to form new memories is ubiquitous across different types of material (e.g., verbal vs. visuospatial) and different modes of presentation (e.g., visual or auditory). Amnesic patients typically also have some difficulty recalling information acquired prior to the onset of injury (retrograde amnesia), but the temporal extent of retrograde memory loss is quite variable.

One of the most striking case reports of amnesia is that of patient H. M. As an experimental treatment for intractable epilepsy, H. M. underwent surgery that removed his medial temporal lobes (MTLs) bilaterally at the age of 27. Although the surgery was largely successful in relieving H. M.’s epileptic seizures, it left him with a profound amnesia, which H. M. described as “like waking from a dream . . . every day is alone in itself.” His anterograde memory loss was apparent both in his daily life and in laboratory testing. For instance, H. M. could not remember where he lived or what he ate for breakfast earlier that day. In laboratory testing, he failed to recall anything about a drawing that he copied merely minutes earlier. This pattern of forgetting was observed across different types of materials and presentation formats (e.g., spoken words, pictures, abstract patterns, spatial locations).

H. M. also experienced retrograde memory loss that extended years prior to his surgery.

It is important to distinguish the memory loss experienced by amnesic individuals and that experienced by patients with more global cognitive decline (e.g., dementia). Whereas amnesics’ cognitive impairments are memory-specific and fairly stable, dementia patients’ cognitive decline tends to be global and progressive.

Etiologies and Neuroanatomical Correlates

Amnesia can result from a variety of causes, ranging from psychogenic (e.g., dissociative disorders) to organic. The focus here is on the organic, neurological conditions that are associated with amnesia. Although lesion location (and extent) as well as symptoms vary as a function of etiology, the pattern that emerges is that the MTL and the diencephalon (including the hypothalamus, thalamus, and epithalamus) are crucial to normal memory functions. Additional frontal lobe damage may lead to deficits in executive and strategic components of memory as well. The most common etiologies are highlighted in this section.

Herpes Simplex Encephalitis (HSE)

HSE is a viral infection that results in hemorrhagic lesions in the brain. In the acute phase, HSE patients are typically confused and disoriented, and they may also exhibit other cognitive impairments, including agnosia, aphasia, and amnesia. The course of illness is variable. In some patients, agnosia and aphasia resolve, and patients are left with a profound amnesia, in the absence of other cognitive deficits. Other patients continue to experience multiple cognitive problems, and yet others recover fully. HSE is associated with damage to the MTL, including the amygdala, hippocampus, and adjacent entorhinal, perirhinal, and parahippocampal cortices. In many cases, damage extends laterally to include the anterolateral and inferior temporal cortex; this damage may be linked to aphasic and agnostic symptoms in some patients. Lateral lesions have also been associated with extensive retrograde amnesia. When the lesion extends anteriorly, it may impact the ventromedial frontal cortex, including the insular cortex and basal forebrain. Laterality effects have also been observed, in which patients with damage focused in the left temporal region...
tend to have deficits primarily in verbal memory, and patients with damage focused in the right temporal region tend to have deficits primarily in visual/nonverbal memory.

Anoxia: A Reduction in Oxygen Supply

Anoxia can result from cardiac arrest, respiratory distress, or carbon monoxide poisoning. Although some patients with anoxic injury present with generalized cognitive impairment, a subset of patients present with selective amnesia. The degree of both anterograde and retrograde amnesia is variable, sometimes with retrograde loss spanning years prior to injury. Typically, the severity of memory loss is determined by the extent of MTL damage. The MTL is only one of several brain regions that are particularly susceptible to oxygen deprivation. As a result of reduced oxygen supply, other areas, including the basal ganglia, thalamus, and diffuse cortical areas, may also suffer damage. In the case of more extensive damage, patients tend to present with more generalized cognitive impairments.

Wernicke-Korsakoff Syndrome (WKS)

In the acute phase, patients with WKS present with confusion, oculomotor abnormalities, and ataxia. Once the confusion clears, patients are left with a severe amnesia and poor insight into their memory problems. The main cause of WKS is chronic alcohol abuse, and resultant thiamine deficiency, although other causes of malnutrition have also been observed. Anterograde amnesia is characterized by a high sensitivity to proactive interference, leading to intrusions from previously encountered (yet currently irrelevant) information. Retrograde memory loss is also common and can extend for decades, although recent memories are typically more affected than are remote ones. The dense amnesia experienced by these patients is often exacerbated by executive dysfunction (i.e., problems in the areas of attention and cognitive control). WKS amnesia is thought to be due to damage to the anterior and dorsomedial thalamic nuclei and the mammillary bodies. Additional frontal damage is responsible for the executive deficit.

Cerebral Vascular Accidents

Infarction of the posterior cerebral artery, which perfuses the MTL, is another common cause of amnesia. In addition to having memory problems, patients with posterior cerebral artery lesions often have reading and visual naming impairments. This pattern of deficit is thought to be due to damage extending to the posterior parahippocampus and the collateral isthmus (a pathway that connects the posterior parahippocampus to association cortex).

Thalamic strokes that impact the anterior thalamic nuclei, dorsomedial thalamic nuclei, or the mammillo-thalamic tract also can result in significant memory loss. Severity of anterograde and retrograde amnesia vary depending on lesion location. Unilateral lesions lead to material-specific amnesia. Executive dysfunction is often observed as well.

Anterior Communicating Artery Aneurysms

In the acute phase of anterior communicating artery aneurysm, patients appear confused and disoriented, and confabulation often co-occurs. In the chronic phase, patients tend to have severe recall deficits, although recognition may be preserved in some cases. The amnesia is due to damage to the basal forebrain, which provides important cholinergic inputs to the MTL. Frontal lesions are also common and may lead to executive dysfunction.

Anterograde Amnesia

One of the hallmark features of the amnesic syndrome is the patients’ inability to consciously recollect information that they have encountered after illness onset, and this includes both personally experienced events (i.e., episodic memory) and generic information about the world (i.e., semantic memory). These types of memory are known collectively as declarative memory because they require conscious recounting of information. Many studies of amnesic patients focus on uncovering the processes that underlie this declarative memory impairment.

Traditionally, researchers have conceived of memory in terms of three key elements: encoding (processing of input), storage (consolidating incoming material for long-term storage), and retrieval (reactivating previously stored information). It is unlikely that amnesics’ memory deficit is due to impaired encoding, as patients are able to maintain information in short-term memory. Because amnesics are able to recall some memories acquired years before amnesia onset (e.g., childhood memories),
it seems implausible that their impairments lie entirely in the retrieval process. Thus, it has been suggested that amnesics' memory problems are due to an inability to consolidate information for long-term storage. Both human and animal research has highlighted the crucial role the MTL plays in this process of consolidation. The MTL is not the site of permanent storage of long-term memory (as previously formed memories can still be retrieved), but rather binds together into coherent representations different elements of an event that are processed and stored in the neocortex.

Even among patients with MTL damage, there is considerable heterogeneity in performance, which has been linked to extent of lesion. Patients with extensive lesions involving the hippocampus and surrounding subhippocampal cortices perform poorly on both recall and recognition tasks, but some patients with damage limited to the hippocampus proper have been described as having impaired recall only. This dissociation can be understood with reference to two cognitive processes that can support declarative memory: recollection, an effortful process involved in the recovery of episodic information; and familiarity, a subjective feeling that results when information is processed fluently. Recollection is critical for recall, but recognition can be supported by either recollection or familiarity. Thus, impairment in recall but not in recognition is thought to reflect selective impairment in recollection. Both human and animal research suggests that the hippocampus proper may mediate recollection, whereas extra-hippocampal MTL regions such as perirhinal cortex may mediate familiarity. This claim remains controversial, however, because some amnesic patients with damage limited to the hippocampus have shown impairments in both recall and recognition.

In patients with frontal damage in addition to MTL damage, compromised encoding and retrieval strategies as a result of executive dysfunction further compound the memory impairment. Executive dysfunction may lead to poor elaboration and organization of incoming information and impoverished memory search and monitoring strategies. The latter may manifest behaviorally as high levels of intrusion errors in recall and false alarms in recognition.

Despite their profound impairment in forming declarative memories, amnesic patients are able to form new memories that do not require conscious access to stored information (i.e., nondeclarative memory). For instance, amnesic patients show intact procedural memory (i.e., learning new motor skills across multiple learning opportunities), eyeblink conditioning (i.e., learning to associate a tone with an air puff to the eye), and repetition priming (i.e., increased efficiency at stimulus processing as a result of single prior exposure). These observations suggest that nondeclarative memory is independent of MTL or diencephalic regions.

**Retrograde Amnesia**

Retrograde amnesia can encompass deficits in memory for personally experienced events (autobiographical memory) as well as generic facts about the world (semantic memory) and oneself (personal semantics). As described earlier, a temporal gradient in retrograde memory loss is frequently observed, with remote memories being better preserved than memories acquired shortly before the time of injury. The temporal gradient for semantic memory typically is time limited, but the temporal gradient for autobiographical memory is more variable. In some cases, it extends a few years (e.g., anoxic patients), whereas in other cases it extends several decades (e.g., WKS patients). The observation of relatively limited retrograde memory loss (i.e., years) is generally consistent with the consolidation view of the MTL, which states that the role of the hippocampus in memory is time limited: The hippocampus is important for the initial binding of event features that are stored neocortically, but over time, direct neocortical connections are formed and memories can be accessed without hippocampal involvement.

The finding of extensive autobiographical memory loss (i.e., decades) challenges that notion, as that time frame extends consolidation beyond what is biologically plausible. This has led to an alternative view, called multiple trace theory, which suggests that retrieval of event-specific details continues to be hippocampal dependent, no matter the age of the memory. According to multiple trace theory, each retrieval of a memory leads to the formation of a new hippocampal trace. The relative preservation of remote compared to recent memories reflects the fact that remote memories, which presumably are retrieved more frequently, are represented by more hippocampal traces and thus more resistant to partial hippocampal damage.
In patients with lesions that include the anterolateral temporal cortex (e.g., HSE patients), extensive retrograde memory loss without a temporal gradient can be observed. This pattern is thought to be the result of damage directly to the neocortical sites where event features are stored.

Summary
Memory impairments are commonly observed in patients with neurological insults, and these deficits can affect new learning as well as the retrieval of pre-existing memories. Studies of amnesia have provided tremendous insight into the cognitive and neural architecture of memory. Theoretical advances into the nature of impaired and preserved memory processes in amnesia can guide development of theory-driven rehabilitation, which may lead to improved quality of life for individuals with memory disorders.

Irene P. Kan and Mieke Verfaellie

See also Memory, Neural Basis; Semantic Dementia

Further Readings

ANALOGICAL MAPPING AND REASONING

Reasoning by analogy is regarded as one of the most sophisticated aspects of abstract thinking. Analogies are systematic matches between two separate areas of knowledge that are based on common relationships shared by the two knowledge areas. For example, researcher Dedre Gentner describes a scientific analogy in which the structure of an atom can be better understood if one understands how it relates to aspects of the solar system. In this case, the nucleus of the atom can be mapped onto the sun, as a central object, and the electrons can be mapped onto the planets, as orbiting satellite objects. Analogies have been used to generate new knowledge and understand novel situations based on what is known about past situations. Successful use of an analogy requires several steps. These steps include retrieving a source analog (the area of knowledge that is known), mapping this onto a target analog (the new area of knowledge), and inferring new information about the target analog based on what is known about the source analog. There are various special requirements, called constraints, which must be placed on analogies in order for them to be used successfully. Several researchers have developed computational models to better understand analogical mapping and inference processes. Recent research has also focused on understanding how analogical thinking operates in the human brain.

Steps in Using an Analogy
Analogies range from relatively simple four-term cases (of the form A is to B as C is to D) to very complex analogies, such as those about political strategies, historical periods, or wars. Regardless of the complexity of the analogy, there are three steps that must occur: retrieval, mapping, and inference.

Analogical Retrieval
Retrieval of a source analog is the first step necessary to make use of prior information in an analogy. This step occurs when an individual encounters a new situation and wishes to predict unknown information about how it behaves or operates. The new situation may be similar to some event or situation that has been encountered and understood previously. The ability to retrieve or access a relevant prior situation, or source analog, from memory is determined primarily by its similarity to the new situation, the target analog. For successful retrieval, the basis for the similarity should be relational, meaning that relationships among elements should be similar across the situations. This can be contrasted with featural similarity, in which the situations look
perceptually similar but operate differently at a relational level. Featural similarity can be distracting when one attempts to retrieve a source analog, as possible source analogs may initially appear appropriate but ultimately, on closer inspection, share little relational similarity.

**Analogical Mapping**

The mapping process occurs when the source and target analogs are aligned and compared. Similar relations are matched between the two domains, and shared relations are paired together. For example, in the solar system-to-atom analogy, relations such as “revolves around” will be paired together so that the satellite objects (planets and electrons) and central objects (sun and nucleus) will be placed in correspondence. Once the mapping process is completed, the individual will have generated an understanding of the overlap that exists between the two domains.

**Analogical Inference**

The final step in the analogy is to generate possible inferences about the new target analog based on what is known about the prior source analog. For example, the knowledge that the planets follow their own paths around the sun may lead to the inference that electrons will follow their own paths around the nucleus. Analogical inferences can then be tested for validity with further analysis of the target domain. It is not guaranteed that analogical inferences will be correct, as their validity depends on the quality of fit between the two domains involved in the analogy.

**Models of Analogy**

Research on analogical thinking has focused on better understanding the process through building theoretical models. These models may then be simulated using computer programs to see how they would solve analogy problems. Most of these models have required that constraints be applied to analogies to enable successful computer simulations. Constraints are pressures that are applied to ensure that valid inferences are drawn. Analogical constraints include the idea that analogies should be based on relations among items in the two domains rather than similar featural appearance. Another useful constraint is that one element in a target analog should match only one element in a source analog. Modeling efforts have been useful in generating predictions about analogical reasoning that can then be tested with human subjects in the laboratory setting.

**Analogy and the Brain**

Recent research on analogy has focused on gaining a better understanding of how analogical reasoning occurs in the brain. Some of these studies have focused on understanding how damage to brain areas affects analogical reasoning. Studies of patients with frontal lobe damage have demonstrated that this area of the brain is important in analogical mapping and in appropriately discounting inappropriate possible matches that share only featural similarity with the target analog. Other studies have used brain imaging methods to demonstrate that the frontal cortex becomes active during analogical mapping compared to other types of comparisons. This area of research indicates that the frontal lobes are important for analogical mapping and inference.

Daniel C. Krawczyk

See also Analogical Reasoning, Models of Development; Similarity

**Further Readings**


prior understanding of the world. Central in adult cognition, analogy is also important for children's capacity to transfer learning across domains and for schema abstraction. Whereas there is general agreement that analogy is important for cognitive development, there is considerable disagreement on the mechanisms underlying children's development of mature, adult-like analogical reasoning. This entry briefly surveys the dominant theories of the development of analogy and then discusses computational models attempting to test these theories.

**Developmental Change in Analogy**

Whereas older children frequently use relational similarity in the service of solving problems, young children typically favor concrete, less relationally complex analogies based on featural similarity. Hypotheses for explaining these differences have centered on changes in relational knowledge and maturation of executive functions.

**Relational Knowledge**

Usha Goswami has argued that children are able to map relations in a rudimentary manner from early infancy, but their later analogical reasoning skills build on prerequisite content knowledge. Thus, children's analogical reasoning becomes more and more adult-like on a domain-by-domain basis as knowledge develops.

Similarly, Dedre Gentner and colleagues hypothesized a “relational shift” during cognitive development such that, as children build knowledge in a domain, they move from attending to similarity based on object features to relational similarity. These authors postulate this process is not an age-related phenomenon but rather is tied to knowledge acquisition. Robert Morrison and colleagues have alternatively argued that the relational shift can be understood as a deficit in inhibitory control in working memory, one aspect of executive functions.

**Executive Functions**

Even when young children can demonstrate relational knowledge in a domain, they frequently have difficulty using analogies requiring integration of multiple relations. Graeme Halford has proposed a theory of *relational complexity* to categorize relations by the number of sources of variation that must be processed in parallel. Halford suggested that on average, children's working-memory capacity is such that after age 2, children can process binary relations, and after age 5, they can process ternary relations. Thus, children of age 2 can perform very simple analogy problems but not problems that require integrating multiple relations.

**Computational Models of Analogy**

Over the past 40 years, many computational models of analogical reasoning have emerged but relatively few attempts have been made to use these models to describe the development of analogy. Efforts to do this can essentially be divided into two branches: (a) efforts to model how children develop relational representations of knowledge, and (b) efforts to model how children use those representations in the service of analogy.

**Building Relational Representations**

All successful models of analogical reasoning operate on structured representations in long-term memory; however, until recently, no explanation existed for how these structured representations might arise. The lack of an account of where relational representations come from has led Robert Leech and colleagues to postulate that analogy might not rely on structured representations but rather on simple associations. However, these approaches have failed to provide an explanation of how children can process progressively more relationally complex analogies or exhibit the flexibility in relational thinking characteristic of adults. A more viable option is that humans can learn structured relational representations from unstructured examples. These representations might then be used in traditional symbolic or symbolic-connectionist models to perform analogical reasoning. Leonidas Doumas and colleagues recently described one such approach that uses comparison to bootstrap learning structured relational representations starting with simple distributed representations of objects as feature vectors.

**Developing Analogical Reasoning**

Assuming a mechanism to learn relational knowledge, a computational model of the development of analogical reasoning must account for why children show a relational shift, why the relational shift is domain specific, and why children initially have difficulty processing relationally complex analogies. In
an effort to capture the relational shift, Gentner and colleagues hand-coded different relational representations in the structure-mapping engine (SME). SME showed a mapping advantage for the representation containing higher order relations similar to that observed in older versus younger children.

An alternative solution to this problem involves assuming that the development of analogy is at least partly tied to maturation of brain systems, particularly areas of the prefrontal cortex known to develop well into adolescence. Previous neuropsychological and neuroimaging analogy studies in both children and adults have shown these areas of the brain to be critical for analogy performance. Working under this assumption, Morrison and colleagues used John Hummel and Keith Holyoak’s LISA (learning and inference with schemas and analogies), a symbolic connectionist model of analogy, to successfully simulate both the relational shift as well as relational complexity effects in children from the age of 3 to 13. All that was necessary to capture these phenomena was a change in LISA’s inhibition parameter, a parameter closely tied to one likely role of the prefrontal cortex. This approach can also benefit from domain-specific changes in representation, when relational representations are chunked presumably as a result of knowledge accretion or expertise. Thus, as relational knowledge develops in a domain, processing can become easier for LISA. This approach provides an explanation for both quick changes in analogy performance as a result of learning and also slower maturational changes as a result of changes in biology.

Robert G. Morrison

See also Analogical Mapping and Reasoning; Similarity

Further Readings


**ANCHORING**

The last two digits of your telephone number have absolutely nothing to do with your life expectancy. And yet, thinking about those two digits may influence your estimates of the amount of time you have to live. First consider, for instance, whether your life expectancy is longer or shorter than those last two digits, and now estimate your actual life expectancy. If your last two digits are small numbers (e.g., 11), you are likely to estimate that your life expectancy is shorter than if those two digits are large numbers (e.g., 99). This is an example of anchoring, a general term used to describe cases in which a person’s judgment or evaluation is influenced by—or anchored on—salient information in one’s environment. In most cases, anchors exert a drag on the judgment and render final estimates biased in the direction of the original anchor value, whether those anchor values are relevant to the judgment at hand or not.

Anchoring has been used to describe two very different aspects of judgment and evaluation. The first is as a phenomenon, to describe cases in which a person’s judgment or evaluation is influenced by an anchor value. The second is as a process, to describe the psychological mechanisms that enable people to make judgments and evaluations under conditions of uncertainty.
Anchoring as a Phenomenon

Psychologists Amos Tversky and Daniel Kahneman conducted a simple, and now classic, experiment in which participants watched an experimenter spin a wheel that landed randomly on a large or small number. People were then asked to estimate whether the percentage of African countries in the United Nations was larger or smaller than that number, and then to estimate the actual percentage. Even though completely random, the anchor value generated by the wheel biased people’s judgments in the direction of the anchor value. For instance, the median estimates of those who considered 10 and 65 as anchor values were 25 and 45, respectively.

Anchoring effects like these have since been demonstrated in a wide variety of contexts. Shoppers buy more Snickers bars when the store’s advertisement recommends buying “18 for your freezer” than when it recommends buying “some for your freezer.” Auto mechanics estimate that repairs will be less expensive when a car owner asks if the cost will be more or less than a relatively low amount than when the owner asks about a relatively high amount. And court judges recommend a longer criminal sentence after considering a prosecutor’s high anchor value than after considering a defense attorney’s low anchor value. Although nonexperts tend to be more strongly influenced by arbitrary anchors, even experts ranging from lawyers to doctors to real estate agents can be influenced in their field of expertise as well.

Psychologists do not believe that anchoring effects are restricted to purely numeric estimates. For instance, anchoring has been used to describe egocentric biases in which people’s estimates of others’ preferences seem to be anchored on one’s own preferences. A person who likes diet cola is more likely to believe that others like diet cola than is a person who dislikes diet cola. So, too, do people’s final impressions of others tend to be biased, or anchored, on their first impressions, even when those first impressions are known to be false or misleading.

Anchoring as a Process

It is easier to demonstrate anchoring effects than to explain why they occur. The problem is that anchoring may be produced by at least three processes: insufficient adjustment, biased accessibility, and numeric priming. Trying to understand when these different processes operate, or whether there are truly different processes at all, is an ongoing topic of research and debate.

Insufficient Adjustment

Tversky and Kahneman suggested that people routinely answer judgments under uncertainty by starting from an initial anchor value and adjusting it to yield a final estimate. A person can estimate the likely cost of a new Toyota Camry, for instance, by adjusting up from the cost of last year’s model to accommodate inflation. Anchoring effects emerge on this account because the adjustments that people make from these anchors tend, for a variety of reasons, to be insufficient.

Biased Accessibility

Asking whether your life expectancy is more or less than the last two digits of your phone number requires you to directly compare your life expectancy with that number. That comparison process is likely to render information consistent with that number selectively accessible. Considering whether your life expectancy is more or less than 22, for instance, is likely to lead you to think about reasons why you might die relatively early in life, whereas considering whether your life expectancy is more or less than 99 is likely to lead you to think about reasons why you might die relatively late in life. Once rendered accessible, that information biased by the anchor value is likely to be used to generate your final estimate and, therefore, to bias judgment.

Numeric Priming

When making numeric estimates, simply being shown a large number is likely to make it easier to recall a large number from memory compared to being shown a small number. This “priming” can influence numeric estimates by influencing information recall. For instance, people think the average runway is longer after considering an anchor value listed in feet compared to in yards.

Nicholas Epley

See also Availability Heuristic; Belief and Judgment; Representativeness Heuristic

Further Readings

Anesthesia and Awareness

The purpose of anesthesia is to render a surgical patient insensitive to pain. Regional and local anesthesia act on the specific sensory nerves and tracts of the skeletal nervous system and spinal cord, leaving the patient conscious and able to communicate and cooperate with the surgical team. General anesthesia, by contrast, operates at the level of the central nervous system, abolishing pain by rendering the patient totally unconscious, in a sort of “controlled coma.” For this reason, research on general anesthesia provides an avenue for investigating fundamental aspects of consciousness. This entry provides a brief overview of modern anesthetic technique and the monitoring of anesthetic depth, and discusses issues pertaining to surgical awareness and memory.

Anesthetic Technique

The first successful use of ether by William Morton on October 16, 1846, is still celebrated as “Ether Day” by anesthesiologists worldwide. Chloroform was introduced in 1847 and gained popularity when used by Queen Victoria during the birth of her eighth child. Modern anesthetic practice, known as balanced anesthesia, employs a “cocktail” of different drugs to achieve three different end points: sedation, loss of consciousness, and muscle relaxation. Commonly employed sedatives include diazepam and midazolam. Typical anesthetic agents include inhalants such as nitrous oxide and oxygen, isoflurane, or sevoflurane, and intravenous drugs such as sufentanil or propofol. Muscle relaxation is typically achieved with a drug such as tubocurarine. Because the neuromuscular blockade induces a total paralysis of the skeletal musculature, the patient must be artificially respirated during the procedure, until the blockade is reversed by a drug such as neostigmine. For the same reason, the patient is unable to communicate with the surgical team or respond behaviorally to requests; this situation raises the question of how anyone knows that the patient is really unconscious.

Monitoring Anesthetic Depth

Clinically, the success of general anesthesia is shown by the patient’s lack of response to verbal commands or to what is euphemistically referred to as “surgical stimulation.” Moreover, postoperatively, the patient will report no sensation of pain during the operation; nor will the patient recall any events that took place during the operation. By these standards, far less than 1% of surgical patients report any surgical awareness. Still, a postoperative interview is a little late to determine that a patient was not adequately anesthetized, and so considerable effort has been made to develop means of monitoring the patient’s state of consciousness, as well as vital functions, during anesthesia. One “on line” measure, based on presumed autonomic signs of conscious awareness, is the PRST score, which takes into account the patient’s blood pressure, heart rate, sweating, and secretion of tears.

Another standard simply relies on biochemical measures of anesthetic concentration in the lungs or bloodstream. A common measure is MAC, which is the minimum alveolar concentration of inhalant anesthetic, measured through the respirator, which eliminates motor response to surgical stimulation in 50% of patients. A weaker concentration, MAC-aware, also known as MAC-awake, typically about 0.3 to 0.5 MAC, is sufficient to eliminate awareness (measured by response to a verbal request), without necessarily eliminating all reflexive motor responses. A stronger concentration, MAC-BAR (1.7–2.0 MAC) is required to block autonomic as well as skeletal reflexes. Including muscle relaxants in balanced anesthesia, then, allows administration of lower doses of anesthetic agents. Similar standards, based on blood plasma levels, have been determined for intravenous anesthetic agents. Because the operational definition of MAC means that 50% of patients will respond to surgical events,
Anesthesiologists generally administer a dose equivalent to approximately 1.3 MAC to ensure adequate anesthesia.

Most modern methods for monitoring the depth of anesthesia involve measures of central nervous system function, such as the event-related potential (ERP, or simply EP) observed in the electroencephalogram (EEG), elicited by somatosensory or auditory stimulation (ERPs can be elicited by visual stimulation, too, but surgical patients’ eyes are closed). The ERP consists of three major components: early (appearing in the first 10 msec after the stimulus), reflecting brain stem activity; middle (10–100 msec), reflecting subcortical activity; and late (100–1000 msec), reflecting cortical activity. As a general rule, adequate anesthesia reduces the amplitude of the various peaks and troughs in the ERP subcomponents, as well as their latency with respect to the stimulus. A frequently employed AEP index of consciousness reflects the degree to which three midlatency subcomponents of the auditory ERP are delayed with respect to normal.

Analyses of the EEG power spectrum (derived by a “fast Fourier transform” of the raw EEG signal) show that adequately anesthetized patients typically have a median EEG frequency of 2 to 3 Hz or less (corresponding to the “delta” activity observed in slow-wave sleep), with a spectral edge frequency at the very high end of the EEG frequency distribution, within or below 8 to 12 Hz (alpha activity). Another derivative of the raw EEG is provided by bispectral analysis (BIS), a popular monitoring technique based on a complicated and proprietary set of transformations based on the amount of high-frequency activity (14–30 Hz), synchronization of the EEG at low frequencies, and the presence of a “flat line” EEG. BIS ranges from close to 100 in patients who are normally awake to values well under 60 in patients who are adequately anesthetized by clinical criteria. In one study, a BIS score of 86 was associated with a 50% reduction in recall of material presented during anesthesia and a score of 64 with a 95% reduction.

Awareness and Memory
Adequate general anesthesia abolishes conscious recollection of surgical events by definition, raising the question of whether there is any unconscious perception of these events, outside of conscious awareness but nonetheless encoded in memory. This question has been addressed by looking for evidence of priming, which occurs when processing of one stimulus facilitates the processing of a later stimulus (in negative priming, the first stimulus inhibits processing of the second). For example, subjects who have recently read the word assassin are more likely to successfully complete the fragment a__a__in with an English word than those who have not.

In fact, some research has found evidence of spared priming for information presented during general anesthesia—at least with some anesthetic agents, and especially at BIS levels above 60. BIS levels below 60 appear to abolish both explicit and implicit memory for surgical events. But even when implicit memory is spared, the effect appears to be limited to repetition priming rather than semantic or conceptual priming. Because the patient is presumably unconscious of surgical events at the time they occur, spared priming, usually construed as an expression of implicit memory, is, in this case, better interpreted as an expression of implicit or unconscious perception. This implicit perception, when it occurs, appears to be analytically limited, in that it permits the analysis of the sound but not the meaning of the stimulus. For that reason, preserved priming during general anesthesia does not justify administering therapeutic suggestions intended to facilitate the patient’s postoperative recovery.

Even when there is some sparing of implicit perception, that does not mean that the anesthetized patient is consciously aware of surgical events as they occur. Still, it would be useful to have an “online” index of conscious awareness that does not depend on inferences from physiology. One possibility is suggested by the isolated forearm technique, in which an inflated blood-pressure cuff is used to restrict the flow of blood to the patient’s forearm while the muscle relaxant is being administered. As a result, the patient is free to communicate manually with the anesthetist. One study of awareness during cesarean section found that although patients might respond to instructions delivered during the first minute or so, response quickly drops to zero thereafter, indicating that the patient is, after all, unaware of what is taking place.

John F. Kihlstrom and Randall C. Cork

See also Consciousness and the Unconscious; Electrophysiological Studies of Mind; Implicit Memory; Neural Correlates of Consciousness; Unconscious Perception
Further Readings


**ANOMALOUS MONISM**

Anomalous monism (henceforth AM) was proposed by the philosopher Donald Davidson and holds that individual mental events are identical to individual physical events. But AM denies that psychological concepts may be reduced to, or analyzed in terms of, physical concepts, however complicated. AM is a monistic theory: In contrast to dualistic theories, it holds that what exists does not comprise more than one sort of thing. But Davidson has rejected the classification of AM as materialist—in his view, there is no basis for saying that everything is “really” or “fundamentally” physical. The following three sections will outline the main claims of AM, survey some of Davidson’s reasons for the irreducibility of the mental, and briefly relate an objection to AM.

**The Theory**

Central to the AM are (a) the claim that, in addition to other particulars like objects, there also exist individual events (an event, in this sense, only occurs once, although the same type of event may recur), (b) a view of causation as a relation holding between individual events no matter how they are described (hence, an extensional relation) but always involving strict laws covering cause and effect, and (c) the distinction between events and descriptions true of them. In Davidson’s view, events are mental insofar as they have true descriptions essentially couched in predicates expressing propositional attitudes (like “believes,” “wants,” etc.); events are physical insofar as they have true descriptions using only physical terms essentially.

AM aims to reconcile three principles seemingly in tension: the principle of causal interaction, the principle of the nomological character of causality, and the principle of the anomalousness of the mental.

1. **Principle of causal interaction**: Some mental events interact causally with some physical events.
2. **Nomological character of causality**: Events related as cause and effect fall under strict (exceptionless) laws.
3. **Anomalousness of the mental**: There are no strict laws on the basis of which mental events may be predicted or explained.

According to Davidson, these are all true. The key to reconciling the anomalousness of the mental with the nomological character of causality and the principle of causal interaction between the mental and the physical, which together suggest its falsity, is to allow that mental events interacting causally fall under strict laws but that these laws can only be stated using nonmental descriptions. On Davidson’s view, only physical descriptions are apt for expressing strict laws. Because, by the principle of causal interaction and the nomological character of causality, mental events do fall under strict laws, they must have physical descriptions. Hence, they are physical.

In AM, the existence of events must be clearly separated from how they are described for purposes of particular explanatory interests. In describing a certain event as an action, and explaining it by a belief and a desire of the agent, together constituting her reason for the action, we employ a vocabulary suited to displaying the action as rational (from the agent’s point of view). But explanatory force also requires that the reason cause the effect, and, in felicitous cases, this is true. We thereby pick out events related as cause and effect; these also have (some, complicated) physical descriptions, but mental events are identifiable as instances of salient general categories only by means of descriptions unsuited to incorporation in strict laws.

**The “Lawlessness” of the Mental**

The claim that the mental is anomalous is based on several interlocking considerations: (a) Mental predicates are true by virtue of being ascribed to an
agent by an interpretation of her verbal and nonverbal behavior. (b) Such interpretation is guided by the overall aim of making the interpretee rational and coherent. (c) Such interpretation is subject to retroactive revision over time: The agent's future behavior may warrant a change in what mental predicates are ascribed to her in the past.

Davidson argues that achieving coherence among the interpretee’s beliefs requires ascribing mostly beliefs that are true by the interpreter’s lights—the so-called principle of charity. Ascription of false beliefs takes special motivation, such as felicitous epistemic circumstances on the interpretee’s part (she couldn’t have known that the chair was repainted and, hence, falsely believes it is still green), or simplification of the overall interpretation.

Completely different considerations guide the ascription of physical predicates. Thus, there is a principled chasm between the criteria for applying mental and physical predicates that blocks analysis of the former in terms of the latter. Whatever coextensionality obtains between mental and physical predicates is accidental, and not law-like. Furthermore, the mental does not constitute a causally closed system: Regularities involving mental types of events will have exceptions due to (actual or possible) intervening physical factors. For example, the putative strict law “whoever wants a glass of milk and believes that she can get one by opening the fridge will open the fridge” may fail to hold in a particular case because someone who has the desire and belief mentioned sprains her ankle en route to the refrigerator. And even if such a regularity would happen to be true, it wouldn’t be a law; it would only be accidentally true. Thus, there cannot be strict laws involving mental predicates. There are, however, non-strict generalizations comprising our folk psychology, which underwrite the singular causal claims implicit in action explanation. For instance, the weaker generalization “if someone wants a glass of milk and believes that she can get one by opening the fridge, she will tend to open the fridge” may well be true, because it explicitly allows for the exceptions blocking strict laws.

The Causal Efficacy of the Mental

Davidson’s interpretation-based view of mentality, in conjunction with AM, suggests that a given mental event could have had different true mental descriptions under different circumstances relevant to interpretation: In some contexts, the event would have been appropriately described as a desire for milk; in others, as a desire for orange juice. Thus, it could have had a different content yet remained the same event. Recall that, for Davidson, we must distinguish the question whether an event exists from how it is described (e.g., as a desire for milk or as a desire for orange juice); the identity of the event does not turn on how it is described.

This has provoked the objection from, among others, Jaegwon Kim that although mental events are causally efficacious on AM, mental properties are not. They may become epiphenomenal, without any effects whatsoever, because the anomalousness of the mental opens the possibility that another, or no, mental description were true of an event although its physical description stays constant, thereby ensuring instantiation of physical law and, thus, causal efficacy.

Davidson has sought to rebut the objection by appeal to supervenience. Informally, the idea is that the mental supervenes on the physical in that “there cannot be two events alike in all physical respects but differing in some mental respect” (Davidson 1980, 214). Davidson argues that “supervenience ensures that mental properties do make a difference to what mental events cause” (Davidson 1993, 15). Having a mental property may be taken here simply to have a true mental description. Somewhat more formally, mental descriptions supervene on physical descriptions in this sense: “A predicate $p$ supervenes on a set of predicates $S$ if for every pair of objects such that $p$ is true of one and not of the other there is a predicate in $S$ that is true of one and not of the other” (Davidson, 1985, p. 242). Here, the predicate $p$ is best thought of as some mental description, whereas $S$ may be thought of as a set of physical descriptions.

The question, however, is whether the pairs of events (or objects) are taken to belong to the same or different possible worlds (ways the world could be). Are we just comparing distinct events in the same situation, in which case supervenience says there must be some physical difference between the events whenever there is some mental difference—or are we also comparing an event with itself in a different possible situation where another mental description may be true of it? It is controversial whether the former reading substantiates the claim that mental properties make a causal difference. And the latter reading appears to be in tension with the principle...
of the anomalousness of the mental, for then the requirement is that there be a physical difference in the event itself whenever there is a mental difference; hence, there will be a law-like, necessary connection between the event's physical properties and its mental properties. This tension may be resolved, however, if the event's physical properties are understood in a very broad sense encompassing all circumstances relevant to interpretation of the subject in which the event takes place.

Sören Häggqvist

See also Emergence; Folk Psychology; Mental Causation; Mind-Body Problem; Philosophy of Action; Physicalism

Further Readings


ANOSOGNOSIA

Anosognosia means “without knowledge of disease.” Joseph Babinski introduced the term in 1914 to describe apparent unawareness of hemiplegia (paralysis of one side of the body) following acute brain lesion. It currently refers to lack of awareness of any deficit resulting from brain damage. This entry summarizes general features, clinical importance, and diagnostic criteria of anosognosia. It reviews (a) differences in “unawareness” and difficulties in use of the term and (b) diverse manifestations and accounts of anosognosia that indicate it is not a unitary disorder. Behavioral dissociations are described exemplifying the specificity, extension, and partiality of anosognosia. Phenomena shown by subsets of patients are cited indicating that unawareness of deficit may depend on patients’ viewpoints and be similar to delusional states in these particular patients. The entry emphasizes the diversity of anosognosia, especially in different neuropathologies.

Occurrence, Importance, Criteria

General Features

Unawareness or denial of deficit can occur for disorders or losses of perception, sensation, movement, action, memory, language, speech, planning, attention, conceptual understanding, musical ability, and emotion. Anosognosia can happen with both focal and diffuse brain damage: stroke, traumatic brain injury, progressive dementias, and tumor. Although more associated with right brain damage, no clear link of anosognosia with specific damage has been established. Following stroke or traumatic brain injury, it usually occurs in the immediate acute phase and remits, but it can last for months, even years. If anosognosia recovers abruptly, sudden awareness of serious deficit can have extreme emotional effects. With progressive brain disease, initial small deficits are usually unnoticed; as they worsen, they are noticed, but as deficits further deteriorate, patients become and remain unaware of them. Unawareness of certain deficits (e.g., a small visual field loss) is unsurprising, especially if their severity progresses slowly. With certain memory or planning losses, patients might not realize their ability has changed. However, unawareness of a sudden dramatic deficit (e.g., blindness, deafness, paralysis) is puzzling and represents a loss of self-awareness.

Clinical Importance and Remediability

Anosognosia is clinically important. While patients are unaware of, or deny, a deficit, rehabilitation is difficult and ineffective, and patients unaware of perceptual or motor deficits can injure themselves. Interventions to improve awareness have minimal or short-lived effectiveness. Occasionally, after physical examination or testing, patients reduce their ratings of strength in plegic (paralyzed) limbs or ability in other deficits, but they quickly return to gross overestimates.
Anosognosia

Diagnostic Criteria

The usual diagnostic criterion for anosognosia is unawareness or denial in conversation, or behavioral evidence. Conversational evidence is a major problem when brain damage affects language or speech. However, often patients show unawareness or denial conversationally but demonstrate awareness behaviorally (e.g., never attempting bilateral actions when hemiplegic) or vice versa (complaining of paralysis but attempting actions requiring affected limbs). Frequently, especially as anosognosia remits, patients know they have a particular deficit, but are unaware of its consequences in specific incapacities; such patients may not be classified as anosognosic.

Differences in Characterization and Explanation

Differences in Meanings and Phenomena

There is little consensus on the characterization, nature, or explanation of anosognosia. One reason is that “unawareness of deficit” has several meanings and refers to different phenomena. (a) Unaware of can simply mean “not knowing”; in technical usage, awareness usually refers to consciousness, and unawareness allows for possible nonconscious knowledge. (b) Awareness as referring to concurrent conscious experience of some state or event is different from generic knowledge as assessed by questions like “How well can you currently do X?” (c) Furthermore, lack of awareness of deficit may differ from a positive belief or assertion of normality; some patients spontaneously deny their disability, whereas others exhibit unawareness only if questioned.

Different Accounts (and Non-Unitariness)

Indeed, anosognosia may have diverse forms and causes and may differ according to the primary deficit. Differences in anosognosia after acute and in progressive brain damage have encouraged separate consideration.

In Acute Brain Damage

In acute brain damage, evidence suggests that anosognosia for some deficits—plegia, blindness—is not a single phenomenon. Consider anosognosia for hemiplegia, common after stroke. Numerous accounts have been offered, and several seem to be true for various cases: loss of proprioceptive information or awareness; a deficit of spatial attention; failure to update bodily knowledge; illusory experience of intended limb movement; refusal to acknowledge movement failure or deficit for personal or social reasons; a state of confusion; lack of concern; conscious unawareness with nonconscious knowledge. Apparent unawareness may in fact be different phenomena in different patients or multiple phenomena in a single patient. Where the deficit is conspicuous, as with hemiplegia, two-factor theories have been proposed: for example, sensory loss plus intellectual loss or confusion preventing discovery of the deficit by observation or inference. Certain explanations fit some deficits, but are implausible for others. Lack of feedback might underlie unawareness of output or action deficits, but it is unlikely to explain unawareness of perceptual or comprehension deficits. Yet, even for deficits of action, concurrent unawareness of specific action failures does not account for the generic unawareness of a deficit in many patients whose repeated failed actions have left intentions unfulfilled and caused injury and embarrassment (e.g., denial of movement problems some time after repeatedly falling out of bed when trying to go to the toilet).

In Progressive Disease

Accounts of anosognosia in dementias also vary (and may apply to other types of brain damage). A progressive deficit in self-monitoring characterizes the changes in encephalitis. Increasing confusion or loss of self-awareness is plausible in Alzheimer’s disease because the pathology spreads to many brain areas. Unawareness of deficit in a function as intrinsic to the function itself (with more functions deteriorating over time) also makes sense in amnesia and neglect. Patients may not appreciate memory loss if they cannot remember how it was before or have lost the concept of memory. In neglect, patients who fail to attend to one side of space may not notice perceptual absence of that unattended spatial content. However, this account cannot be generally true as many patients with amnesia and with neglect, and most patients with other severe deficits, are aware of their deficiency.

Motivational Accounts (in Acute Brain Damage)

Motivational accounts view anosognosia as a defense. Although refusal to acknowledge illness occurs in many conditions, several studies find no evidence that anosognosic patients tend to deny illness or deficit generally. Motivation may underlie denial in some patients, although a gross mental change may allow motivational features to manifest.
themselves even if they do not cause the unawareness. Elaborate or implausible confabulations that some patients produce negate or rationalize their condition but suggest either something more than psychodynamic repression alone or an attempt to make sense of an uncomprehended situation, disregarding plausibility. Many people who suffer equivalent deficits that are not due to brain damage (e.g., blindness, paralysis) do not deny such deficits. It is unclear why brain damage should produce or legitimize denial or repression.

Dissociations and Complexity

Three Aspects of Anosognosia

Showing Dissociations

Different kinds of dissociation define three aspects of anosognosia. (Dissociation here refers to one task or behavior being deficient or affected while another is not.)

Specificity

Patients with more than one primary deficit are often aware of one while unaware of another; severity or personal importance of deficit does not predict which one the patient is unaware of. Anosognosia in patients with otherwise normal mental abilities suggests it can be due to a restricted cognitive deficit.

Extension

Different kinds of awareness can be compromised. Patients with hemiplegia may admit their plegia but greatly overestimate their motor capacities. Patients with aphasia are often aware of having aphasia but unaware of their errors when speaking. The converse occurs where patients are aware of inability at the time of failure but moments later deny a deficit. Patients with Huntington’s disease are frequently unaware of their motor or memory deficits as measured by questionnaire but can accurately predict their performance on motor or memory tasks.

Partiality

Unawareness of deficit can be less than total, and in different ways. Patients may admit to a more minor deficiency, attributing it to a trivial cause. Many deny severe deficits but are willing to stay in the hospital and be cared for. This may indicate either nonconscious knowledge or split awareness.

Contradictory Dissociations (Dissociative or Delusional State)

Intriguing dissociations that occur in certain subsets of patients illustrate the complexity and non-unitariness of anosognosia. In jargon aphasia, mainly in the acute stage of Wernicke’s aphasia and global aphasia, patients utter neologisms and unintelligible “jargon” and are typically unaware of their speech and comprehension deficits. Some patients rate recordings of their own speech as normal but rate recordings of someone copying their utterances as highly deficient. They are satisfied with their own writing but reject it as meaningless when transcribed by an examiner.

Patients anosognosic for hemiplegia rate their ability on bimanual and bipedal tasks as normal. But a subset of these patients, when asked how well the examiner could perform those activities if he or she were in the same state as the patient, say that the examiner would be unable to do them “because you would need both hands.” When asked how they would do each task (e.g., tie a knot), they do not describe normal procedures (as other patients do) but give bizarre and implausible responses (“My right hand is clever,” “I’d use a machine,” “I’m a good sportsman”). Patients among this group, when asked if their plegic arm is weak, deny it, but if asked “Is this arm ever naughty? Does it ever not do what you want?” they readily assent and may say they will punish that arm for disobeying. These patients seem normal when discussing topics other than their plegic limbs. Their responses indicate they are aware in some way of their plegia, and several features of their behavior suggest their anosognosia is a delusional state (not simply a false belief held against contrary evidence, but rather an anomalous state of mind) manifested only regarding the deficit. Such phenomena suggest awareness may be linked to attentional stance or viewpoint, for example, judgments made in a detached stance or from another’s viewpoint. They are reminiscent of similar findings with psychotic delusional patients. However, these patients are not typical of all with anosognosia.

Final Comment

Many anosognosic phenomena suggest kinds of dysfunction in self-awareness or self-monitoring. Although the Greek term connotes a single medical
condition, current research usefully focuses on anosognosia’s different forms in different pathologies.

Anthony Marcel and Mariia Kaliuzhna

See also Aphasia; Attention and Consciousness; Confabulation; Delusions; Self-Consciousness; Self-Knowledge

Further Readings


ANTI-INDIVIDUALISM ABOUT COGNITION

Anti-individualism about cognition is a view about the relationship between thought and the world. In particular, it holds that how a subject represents the world in cognition sometimes depends on the objects and properties in her environment. This dependence regards the very individuation of her mental representations: How she represents the world in cognition depends, in part, on what worldly objects and features she has interacted with. Interaction here means not merely in the causal sense that she would have been caused to have different thoughts, but in the constitutive sense that she would not have had the thoughts she had if her environment had been suitably different. As a thesis about the individuation of mental states, anti-individualism derived from work in the philosophy of language in the 1970s. This work, for example, by Saul Kripke, Keith Donnellan, and Hilary Putnam, subsequently came to be known as the “new” theory of reference. In the decade following the new theory of reference, several theorists (including Tyler Burge, Gareth Evans, and John McDowell) drew inspiration from this work for the philosophy of mind. The result was anti-individualism about cognition. This entry will present anti-individualism as it contrasts with a broadly Cartesian conception of mental states, will identify the main sorts of argument that have been offered on its behalf, and will discuss one of its salient epistemological implications.

Individualistic and Anti-Individualistic Approaches to Cognition

Anti-individualism about cognition is perhaps best appreciated by considering its rival, a certain Cartesian conception of mental states. Thinking of all of the ways in which his beliefs might turn out to be false, René Descartes (in his Meditations) imagines that he is the victim of an Evil Demon—one whose deceptive powers extend so far as to create the illusion of a world “external” to Descartes’s mind. Descartes reasoned that, if this “Evil Demon Hypothesis” were true, then most of his beliefs about the external world would be false. In reasoning in this way, Descartes assumed that, although the Evil Demon might affect the truth of what he (Descartes) believed to be the case, the Evil Demon could not affect the content of what Descartes believed to be the case. This assumption embodies an intuitive and highly influential conception of the relation between mind and world. On this conception, facts about what a person mentally represents to be the case—whether sub-personally (e.g., in the early stages of visual processing) or else in her person-level representations (e.g., in her fears, hopes, thoughts, beliefs, worries, etc.)—can be determined prior to, and in a way that does not depend on, any inquiry into the nature of the worldly objects and properties she aims to be representing. This conception amounts to an individualistic approach to cognition: It holds that how a subject is representing the world in cognition does not depend on anything “external” to the individual herself.

Anti-individualism about cognition is the denial of this individualistic approach to cognition: It holds that the mind depends for the very contents of its representations on the objects and properties
it aims to represent in cognition. Broadly anti-individualistic views of cognition have been developed in a variety of different theoretical settings. Taking a biological-evolutionary approach to the mind, Ruth Millikan and David Papineau have argued that the human representational system is best construed as representing particular kinds of objects and properties in the external environment of our early hominid ancestors. Somewhat relatedly, Hilary Putnam, Fred Dretske, and Tyler Burge have developed anti-individualistic accounts of thought by appeal to the objective nature of the properties represented in perception. Still other thinkers have argued for anti-individualist conclusions by appeal to the relationship between language and thought. The source of most of these arguments is Putnam’s 1976 development of the hypothesis of the “division of linguistic labor.” According to this hypothesis, words of a public language express the concepts they do in virtue of the systematic tendency among speakers to defer to relevant experts for the proper conceptual explication of these words. For example, although any English speaker can use the term molybdenum to refer to molybdenum (a particular silvery metal), few of us would be able to describe this metal in any detail, and still fewer would be able to distinguish it from all other metals (either by observation, experimentation, or description). The availability of the word molybdenum, as a device for referring to molybdenum, depends on the deference ordinary English speakers would have to the relevant experts—the ones who have the knowledge and ability to recognize the metal and distinguish it from others. On the basis of this phenomenon of “semantic deference,” Tyler Burge argued in 1979 for anti-individualistic conclusions regarding a broad swath of human thought. (See, also, Sanford Goldberg’s work for a variant on this sort of argument.)

Versions of Anti-Individualism

As Jessica Brown has noted, three main versions of anti-individualism can be distinguished: natural kind anti-individualism (NKA), singular thought anti-individualism (STA), and social anti-individualism (SA). We might also note that there is a difference between anti-individualism regarding the meaning of words and anti-individualism regarding the content of thought (and of our mental representations more generally).

NKA is a claim about mental representations of natural kinds, those kinds that make up our natural (as opposed to human-made) world: electrons, water, and oxygen, but not Tuesdays, top hats, or toasters. Inspired by Putnam, NKA holds (roughly) that thoughts regarding natural kinds are dependent for their individuation on the nature of the kinds in question: Differences in the types of entity being represented make for differences in mental representations, even if the representing subject herself is ignorant of the differences between the types of entity. Putnam’s famous example of Twin Earth is supposed to illustrate the point. Imagine two planets that are exactly alike except for one difference: Although both planets have a transparent, tasteless, thirst-quenching liquid that is found in rivers and streams (and which flows in the taps and showerheads), the liquids in question differ in their chemical nature. So, whereas on one of the planets (Earth) this liquid is H₂O, on the other planet (Twin Earth) it is a liquid with a very different chemical nature than H₂O. (Putnam uses XYZ as a stand-in for the complicated chemical nature of the liquid on Twin Earth.) However, on both Earth and Twin Earth, the speakers refer to the liquid in question using the word-form water, and in all other respects, the languages spoken on the two planets are the same. Now imagine a subject who has resided for her life on Earth, and her doppelgänger who has resided for her life on Twin Earth, both currently thinking a thought that each would verbalize as “Water is wet.” Assuming that both are ignorant of the chemical nature of the liquid in question, and that neither could distinguish the watery liquid on Earth from the watery liquid on Twin Earth, even so, according to NKA, these subjects are thinking different thoughts. More specifically, they are representing the world differently: One is representing one particular liquid (H₂O) as wet, and the other subject is representing another particular liquid (XYZ) as wet—and this is so, even though the subjects are ignorant of the difference between the liquids and of the difference between these content-distinct thoughts.

STA is a claim about mental representations of particular objects in our environment: this particular woman, that particular apple. Developed by Gareth Evans and John McDowell, STA regards singular thoughts as distinct when they regard two different objects—even if the subjects themselves are unaware of the fact that different objects are involved. Thus, if
at one time you observe a bird and think “That bird is a goldfinch,” and at some later time you observe a different bird (albeit indistinguishable to you) and think, “That bird is a goldfinch,” then STA regards you as thinking two different thoughts. More specifically, you are representing the world differently in the two thoughts: In one thought you are representing a particular bird, $b_1$, as a goldfinch, and in the other you are representing a different bird, $b_2$, as a goldfinch.

SA, which is owed to Tyler Burge, is a claim about those mental representations whose content depends on the division of linguistic labor. Here another Twin-Earth style argument can be made. Imagine that on Twin Earth the language uses the word molybdenum to refer to aluminum and aluminum to refer to molybdenum. Now imagine doppelgängers on Earth and Twin Earth who are ignorant of the difference between aluminum and molybdenum, both of whom express a thought with the sentence-form “Molybdenum is a metal.” According to SA, they are thinking different thoughts. More specifically, they are representing the world differently: The Earthling is representing molybdenum as a metal, and the Twin Earthling is representing aluminum as a metal.

Epistemic Implications of Anti-Individualism

Anti-individualism about cognition is widely regarded as having striking implications for epistemology (the theory of knowledge). Salient among these are its alleged implications regarding thinkers’ knowledge of their own states of mind. Most philosophers endorse the following doctrine of first-person authority: Without recourse to empirical evidence, each thinker is in a position to know her own current thoughts in a special authoritative way. However, many philosophers have argued that this doctrine is not compatible with anti-individualism about cognition. After all, if how one represents the world in thought depends for its individuation on environmental factors, and if these factors can be known only through ordinary empirical ways, then how can it be that, without recourse to empirical evidence, a subject is in a position to know what she is currently thinking? Although no consensus has been reached, a majority of the philosophers who think about these issues have concluded that anti-individualism is incompatible with first-person authority.

Sanford Goldberg

See also Concepts, Philosophical Issues; Concepts and Language; Descriptive Thought; Extended Mind; Indexical Thought; Object-Dependent Thought

Further Readings


ANXIETY DISORDERS

This entry provides basic information regarding the various anxiety disorders, including characteristics and prevalence rates for each disorder. The etiology or vulnerabilities to develop an anxiety disorder is considered along with various effective treatments that are available. Other issues regarding diagnosing anxiety disorders as well as their impact on society are discussed.

Introduction

Anxiety disorders represent a major global mental health problem and are the most prevalent category of psychological disorders in the United States, with a lifetime prevalence estimate of 28.8%. Similar estimates have been reported in surveys from Canada, Great Britain, Australia, and the Netherlands. The overall economic impact of anxiety disorders in the United States has been estimated
Anxiety Disorders
to be over $42 billion per year, or an annual cost of over $1,500 per individual with an anxiety disorder. Direct psychiatric treatment and pharmaceutical costs represent 31% and 2% of the total costs, respectively. Workplace costs, the majority of which stem from lost productivity, represent 10% of the total cost. Over 50% of the total costs result from direct non-psychiatric medical costs, which are believed to result from medical treatment based on non-diagnosis or improper diagnosis. Thus, most of the annual economic impact of anxiety disorders can be avoided or eliminated by proper diagnosis and by the proliferation of effective and accessible treatment options. Indeed, many treatments with documented efficacy for anxiety disorders exist, including pharmacological and psychotherapeutic options.

Despite the wealth of research on anxiety and anxiety disorders, uncertainty exists regarding where to draw the line between separate diagnoses as well as thresholds for determining clinical severity. Indeed, although evidence strongly suggests that the experience of anxiety and fear is universal, cultural variations in the presentation and prevalence of specific fears highlight the non-universality of current diagnostic classification.

Defining and Diagnosing Anxiety
Fear and anxiety are considered normal emotional or cognitive-affective states that are part of the defensive motivational system. Fear is seen as the response to a perceived immediate threat and is typically characterized by an intense physiological response (fight or flight), focused attention on the perceived threat, and an escape motivation. Anxiety serves as the defensive system that scans for possible or future threats and involves a less pronounced physiological response, increased cognitive activity to detect threat (e.g., hypervigilance, worry), and increased avoidance motivation. Although these are normal responses designed to protect the individual from immediate or future threats, in some individuals the responses can become sufficiently excessive that they cause significant impairment or distress and are termed anxiety disorders.

In the most recent revision of the American Psychiatric Association’s Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR), 12 distinct anxiety disorder diagnoses are codified; this is a dramatic increase over previous editions. Currently, the DSM-IV-TR includes panic disorder without agoraphobia, panic disorder with agoraphobia, agoraphobia without history of panic disorder, specific phobia, social phobia, obsessive-compulsive disorder, acute stress disorder, generalized anxiety disorder, post-traumatic stress disorder, anxiety disorder due to a general medical condition, substance-induced anxiety disorder, and anxiety disorder not otherwise specified. While it is acknowledged that the DSM system is not perfect, particularly for anxiety disorders, many would argue that it provides a sufficient framework.

Panic Disorder and Agoraphobia
Panic disorder is characterized by the unexpected, or seemingly out-of-the-blue, occurrence of two or more panic attacks that are accompanied by at least one month of apprehension regarding the recurrence of further panic symptoms. This disorder affects approximately 3% to 5% of the U.S. population during their lifetime. Avoidance of places or situations that are associated with previous panic attacks or panic symptoms is a common behavioral repercussion, called agoraphobia, and is referred to as panic disorder with agoraphobia. Some individuals, however, may be diagnosed with panic disorder without agoraphobia, in which case they do not avoid places or situations for fear that it will set off another panic attack. Agoraphobia without history of panic disorder is another diagnostic distinction where the person may avoid places or situations because of fears of other feared bodily functions for which they do not feel in control (e.g., losing control of the bladder or rectum) but are not associated with fear of panic symptoms in particular.

Social Anxiety Disorder/Social Phobia
Social anxiety disorder, also referred to as social phobia, is characterized by an intense fear of social situations, meeting new people, or entering situations where they may be observed or evaluated by others. This disorder occurs in approximately 7% or more of the U.S. population. Individuals with social anxiety disorder are overly concerned that they may do something to embarrass or humiliate themselves in front of others. They also often avoid any place or situation where this may occur or, when forced, endure the situation with great distress.
Obsessive-Compulsive Disorder

Obsessive-compulsive disorder (OCD) is characterized by repeated thoughts or behaviors that are difficult to control and are disturbing to the individual. Roughly 2% of adults in the United States are currently diagnosed with this disorder, and double this estimate are diagnosed at some point in their lifetime. Obsessions are thoughts, images, or impulses that repeatedly occur and may seem senseless, such as fear of germs or becoming contaminated; doubts about whether locks have been locked, appliances turned off, or paperwork completed accurately; aggressive impulses regarding hurting self or others; sexual, religious, or satanic thoughts or images; or fear of accidentally harming others. Compulsions are behaviors that are performed repeatedly to help alleviate anxiety and may include counting objects, letters, or numbers; checking repeatedly that locks are locked, appliances are turned off, or any other detail is properly taken care of; washing repeatedly and excessively; hoarding or keeping unnecessary materials such as newspapers or trivial items; repeating phrases or words over and over; or strictly adhering to particular sequences or rules, including the symmetrical alignment of objects, performance of ritualistic acts, or specific routines. These thoughts or behaviors often preoccupy a major portion of these individuals' daily lives, and they often fear that horrible consequences will occur if particular rules or routines are not followed.

Generalized Anxiety Disorder

Generalized anxiety disorder (GAD) is characterized by continual and excessive worry regarding a number of areas of life and is estimated to affect 3% to 4% of the U.S. population. Individuals with GAD often find that they worry about several life domains, including minor matters, work or school, family, finances, interpersonal relationships, health of self and significant others, and community or world affairs. These worries are in excess of what most people experience as intermittent worry about daily life where the worry may seem unreasonable or out of proportion to reality, such as constant worry about interpersonal relationships despite no actual conflict. Additionally, individuals are diagnosed with GAD if these worries have occurred over a 6-month period and have been bothersome over this period more days than not. The constant worry is also accompanied by other physical symptoms that may include restlessness, feeling easily tired, difficulty concentrating, irritability, muscle tension, and difficulty with regular sleep for more days than not over the same time period and that have been disruptive in the individual's life.

Post-Traumatic Stress Disorder

Post-traumatic stress disorder (PTSD) is an anxiety disorder that sometimes develops after a traumatic or life-threatening event (e.g., rape, assault, combat, motor vehicle accident) and is diagnosed 6 months after the event if symptoms persist. Approximately 8% of the U.S. population suffers from this disorder. Many individuals with PTSD experience nightmares of the event, mentally relive the event, have an exaggerated startle response to loud noises or unexpected events, or experience intrusive memories of the event. Individuals may put forth considerable effort to avoid thoughts, feelings, or conversations about the event and avoid any activity, situation, or people that may remind them of the event. Additional symptoms may include memory loss of the event, loss of interest in usual activities, or becoming detached or withdrawn from others. These symptoms are typically quite distressing and disruptive to the individual's life.

Specific Phobia

Specific phobias are a set of disorders that involve extreme fearfulness of particular stimuli, including animals (e.g., bees/insects, spiders, snakes), the natural environment (e.g., heights, storms, water), blood or injections, situational type (e.g., small enclosed places, elevators, air travel), or “other” type, including fear of medical or dental procedures, vomiting, or choking. These disorders surpass the normal type of fear in that they cause significant distress and substantially interfere in one's life.

Other Anxiety-Related Diagnoses

Additional anxiety-related diagnoses in the DSM-IV-TR include anxiety disorder due to a general medical condition (anxiety that is the direct effect of a medical condition) and substance-induced anxiety disorder (anxiety related to substance use, abuse, or withdrawal). The presentation of anxiety that does not meet diagnostic criteria for another anxiety disorder but causes significant interference or distress
for the individual is called anxiety disorder not otherwise specified.

In addition to the high prevalence rates of anxiety disorders, high rates of comorbidity (or co-occurrence with another disorder) exist within the anxiety disorders, as well as between anxiety and depressive disorders. Estimates suggest that patients average 2.1 depressive and anxiety disorder diagnoses, and roughly 50% to 60% of individuals with an anxiety disorder have a comorbid anxiety or mood disorder diagnosis. Given that the development of the anxiety disorder typically precedes the development of the mood or substance use disorder, anxiety difficulties may contribute to further distress and impairment.

**Structure of Anxiety**

Although *DSM-IV-TR* describes the anxiety disorder diagnoses as independent or unique, considerable evidence suggests that they share the same or similar underlying etiological and functional bases. These transdiagnostic (or unified) conceptualizations of anxiety articulate that the anxiety disorders differ only in what initiates or is the focus of the emotional states, as in the possibility of future panic attacks for panic disorder or public speaking for social anxiety disorder. Much of the theoretical underpinnings of transdiagnostic models are based around the constructs of neuroticism, trait anxiety, and negative affectivity. Negative affectivity is a more commonly used term that describes the anxiety response and has been described as a general temperamental sensitivity to negative stimuli resulting in feelings of fear, anxiety, depression, guilt, and self-dissatisfaction. This construct of a generalized vulnerability is similar to Aaron Beck’s cognitive models of danger schemata with a perception of a limited capacity to control or predict negative outcomes because of perceptions of uncontrollability or unpredictability. Seen as a common underlying factor contributing to both anxiety and mood disorders, negative affectivity may explain the high rates of comorbidity and similarity between mood and anxiety disorders.

**Etiology**

A considerable body of evidence suggests a heritable component in clinical anxiety that is common across diagnoses. Multivariate genetic analyses have provided strong support for the idea that genetic transmission does not involve specific anxiety disorders but rather a common nonspecific diathesis toward anxiety and other disorders related to negative affectivity. Researchers have concluded that genetics provide a general predisposition for affective disorders, but environmental factors are largely responsible for determining the specific disorder manifestation. Subsequent studies generally support the notion of a nonspecific genetic contribution to the development of anxiety disorders.

A lack of a developed perception of control at early developmental stages may also play a large role in the development of anxiety. Overprotective parenting can lead to children not developing confidence and self-efficaciousness. If parents or primary caregivers limit a child’s ability to develop perception of control over the environment, the child may lack opportunities to learn how to fail and recover, how to navigate through life on his own, and how to make sense of the world around her. In combination, genetic and early learning experiences promote a generalized vulnerability to the development of anxiety and emotional disorders, with subsequent traumatic conditioning, vicarious observational learning, or information transmission leading to the development of specific fears and diagnoses.

**Treatment**

**Pharmacological Therapy**

Many effective pharmacological treatment options exist for anxiety, including tricyclic and heterocyclic antidepressants, serotonin reuptake inhibitors (SSRIs), monoamine oxidase inhibitors, azapirones, beta-adrenergic blockers, and benzodiazepines. The use of psychotropic medications in treatment can add to improvement for some individuals, usually those with more severity. Anxiolytics disrupt the particular flow of brain circuitry that regulates anxiety. However, these effects appear to be maintained only while the individual is taking the medication, and relapse rates with anxiolytics are high. SSRIs are considered the frontline pharmacologic intervention for anxiety disorder, with meta-analyses suggesting strong efficacy with modest relapse rates and side effects compared with other pharmacologic interventions for anxiety.

**Cognitive Behavioral Therapy**

Encouraging efficacy data exist across the range of cognitive behavioral treatments (CBT) for anxiety-related disorders. Recent meta-analyses of CBT outcome studies support the effectiveness of such
Aphasia treatments across the anxiety disorders and, generally, the superior efficacy of CBT approaches over non-CBT psychosocial treatments. Most CBT protocols include psychoeducation regarding anxiety, followed by instruction in cognitive restructuring techniques that address individuals’ anxious beliefs. Individuals are gradually exposed to their specific fears, beginning with less distressing situations or elements, and are encouraged to use their cognitive restructuring skills to view their anxious beliefs in a more logical and balanced manner. Work done outside of treatment is another critical component of CBT, where these individuals continue to conduct their own exposures throughout the week to habituate them to their fears and eventually eliminate avoidant behaviors. Homework generally also includes self-monitoring of general anxiety, stress, and depression throughout the treatment phase.

**Conclusion**

Anxiety disorders represent one of the most prevalent and impactful classes of psychological disorders. Seen as excessive manifestations of basic emotional and cognitive-affective responses, anxiety disorders are currently divided into 12 distinct diagnoses, although recent arguments have been made that the similarities among the diagnoses outweigh the differences. Although the exact causes of anxiety disorders remain somewhat unclear, research suggests that genetic and early learning influences interact to create a generalized predisposition toward emotional disorders, which can then be activated by subsequent learning or experience. Highly efficacious treatments exist for anxiety disorders, particularly those based on cognitive behavioral therapy and SSRIs.

*Peter J. Norton and Suzanne C. Klenck*

**Further Readings**


**Aphasia**

Aphasia refers broadly to disorders of language, which may be developmental (occurring because of an abnormal course of development) or acquired (occurring because of external disease such as stroke, brain tumor, etc.). The term is used somewhat inconsistently. For instance, the disorganized incoherent speech of schizophrenic patients is usually considered a manifestation of a thought disorder and is not considered an aphasic disorder, even though schizophrenic language often contains nonsense words and ungrammatical sentences that could easily be considered signs of aphasia. These differences in the use of the word *aphasia* are due to historical factors (what type of physicians used which terms) that are unimportant today. What is important is the nature and the neurological causes of language disorders. Knowledge about these topics has changed rapidly in the past three decades because of the contributions made by cognitive psychology and neuroimaging. This entry first presents a widely utilized clinical perspective on aphasia and then reviews modern approaches and results.

**Clinical Aphasic Syndromes**

The clinical approach to aphasia identifies some 7 to 10 aphasic syndromes and classifies aphasic patients
into these syndromes. These syndromes are the following:

1. **Broca's aphasia**, a severe expressive language disturbance reducing the fluency of speech in all tasks (repetition and reading as well as speaking) and affecting elements of language such as grammatical words and morphological endings, without an equally severe disturbance of auditory comprehension

2. **Wernicke's aphasia**, the combination of fluent speech with erroneous choices of the sounds of words (phonemic paraphasias) and an auditory comprehension disturbance

3. Pure motor speech disorders—**anarthria, dysarthria, and apraxia of speech**—output speech disorders due to motor disorders, in which speech is misarticulated but comprehension is preserved

4. **Pure word deafness**, a disorder in which the patient does not recognize spoken words, but spontaneous speech is normal

5. **Transcortical motor aphasia**, in which spontaneous speech is reduced but repetition is intact

6. **Transcortical sensory aphasia**, in which a comprehension disturbance exists without a disturbance of repetition

7. **Conduction aphasia**, a disturbance in spontaneous speech and repetition, consisting of fluent speech that contains many phonemic paraphasias, without a disturbance in auditory comprehension

Other syndromes include **anomia** (an inability to find words), **global aphasia** (affecting all aspects of language use), and **isolation of the speech area** (leading to a global aphasia except for sparing of repetition). There are also disorders affecting written language (**pure alexia**, in which a patient can write but not read), and various disturbances of writing (**agraphia**).

The clinical value of these syndromes largely comes from their relation to the location of lesions. Broca's aphasia, which affects expressive language alone, is due to lesions in Broca's area, the center for motor speech planning adjacent to the motor strip. Wernicke's aphasia follows lesions that disturb the representations of word sounds located in Wernicke's area, the cortex next to the auditory receptive areas. Pure motor speech disorders arise from lesions interrupting the motor pathways from the cortex to the brain stem nuclei that control the articulatory system. These disorders differ from Broca's aphasia because they are not linguistic; they affect articulation itself, not the planning of speech. Pure word deafness follows bilateral lesions cutting Wernicke's area off from the areas of the brain that affect the transmission of sound input into Wernicke's area. Transcortical motor aphasia results from the interruption of input from areas where concepts are formulated to Broca's area. Transcortical sensory aphasia follows lesions between Wernicke's area and areas where concepts are formulated. Finally, conduction aphasia follows from a lesion between Wernicke's area and Broca's area.

The basis for clinical aphasic syndromes consists mostly of the relative preservation of patients' performances in the usual tasks of language use (speaking, understanding spoken language, reading, and writing). This contrasts with other historical approaches to the description of aphasia, which concentrated on the conditions under which language is used (e.g., a patient's motivational state). These approaches have not been productive and have largely been abandoned by clinical aphasiologists. The clinical approach thus established the basic phenomena that are relevant to describing and understanding aphasia. In addition to emphasizing performance on tasks, clinical approaches to aphasia include descriptions of linguistic and psycholinguistic disorders of language. For instance, one feature of Broca's aphasia is **agrammatism**, a pattern of speech that consists primarily of substantive words, that is, nouns, action verbs, or significant modifiers, and short phrases with simple syntactic structure. However, the clinical approach is very limited in characterizing these phenomena fully. This fuller characterization has become the province of modern studies of aphasia.

**Psychological and Linguistic Approaches to Aphasia**

In the past 30 years, researchers have begun to address the limitations of the clinical approach to aphasia by developing detailed models of many aphasic abnormalities in relation to models of linguistic structure and psycholinguistic processing. A brief introduction to linguistics
Aphasia and psycholinguistics is required for readers to understand this work.

The language code connects particular types of representations to particular aspects of meaning. Words connect phonological units with items, actions, properties, and logical connections. Word formation forms words from existing words. Sentences relate words in hierarchical syntactic structures to each other to determine semantic relationships between them, such as who is accomplishing or receiving an action. Discourse relates sentences to express temporal order, causation, inferences, and other semantic values. Depending on the level of detail at which language is described, there may be hundreds of independent language processing impairments. For instance, we may recognize a disturbance of converting the sound waveform into linguistically relevant units of sound (acoustic-to-phonemic conversion), or we may recognize disturbances affecting the ability to recognize subsets of phonemes, such as vowels, consonants, stop consonants, fricatives, nasals, and so on.

The psycholinguistic approach describes aphasic symptoms as disorders of these linguistic representations and psycholinguistic operations. There are far too many descriptions of these disorders to review them all; the following discussion focuses on one that affects an important aspect of language and that also connects with neurological facts—disturbances affecting the meanings of words.

Meanings of words are part of a specialized memory store called semantic memory. Most research has focused on disorders of the meanings of words that refer to objects. Disturbances of word meanings affect the ability of a patient to identify a word or a picture as an example of a specific concept and, therefore, cause poor performance on word-picture matching and naming tasks for affected words. Other, more detailed studies of these disorders suggest that some patients have lost specific items in semantic memory and others cannot retrieve items in specific tasks; that disorders may be specific to words or to pictures; that disturbances may be category specific (they may affect concepts related to living things and foods, man-made objects, abstract or concrete concepts, and nominal or verbal concepts); and that some patients cannot match words to pictures or name objects but show evidence of understanding words unconsciously (they show semantic priming effects, responding more quickly to a word when it has been preceded by a semantically related word). One group of researchers has reported that disturbances affecting word meaning co-occur with disturbances of reading irregularly spelled words but not regularly spelled words, and has suggested that this combination of abnormal performances is due to a disruption of memories for single entities (regularly spelled words can be sounded out from their parts). Studies such as these have led to investigations and models of concepts and word meanings and how they are activated normally. As mentioned earlier, similar studies have been undertaken for many linguistic structures and psycholinguistic operations.

This approach to aphasia does not attempt to classify patients into a small number of syndromes. Instead, it allows patients to have numerous deficits. The advantage of this approach is that it provides more detailed information for professionals who treat patients and for neuroscientists who wish to approach the organization of brain by correlating deficits with lesions. Its relatively short history and its reliance on detailed models of linguistics and psycholinguistics that health care providers are not familiar with have limited its clinical adoption, however.

**Neurological Correlates**

Aside from providing information about what aspects of language are abnormal and might be targeted in therapy, studies of aphasia can be used to diagnose diseases themselves. For instance, a (relatively) isolated disturbance of word meanings and concepts relating to objects is seen in a disorder known as semantic dementia. This condition arises following progressive damage to both anterior temporal lobes, and is associated most commonly with a particular type of pathology (discussed later).

Not all symptoms are so specific; many arise following various types of disease. For instance, primary progressive aphasia (PPA), a set of dementing diseases whose primary initial symptom was aphasia, has been divided into three main types: nonfluent PPA, logopenic progressive PPA, and semantic dementia. Nonfluent PPA is very similar to Broca’s aphasia and logopenic PPA is very similar to a combination of anomia and conduction aphasia. The diagnosis of PPA, stroke, tumor, or other disorders is made more on the basis of the time course of symptom development and the radiological findings than on the basis of the language disorder itself.
Understanding the deficits in aphasia is also important in using clinical-pathological correlations to help understand the neural organization that supports language. As mentioned earlier, the clinical syndromes have value for determining the site of brain damage, and these correlations have led to models of brain organization. These models are being supplanted by models based on more detailed descriptions of deficits, functional neuroimaging, neuropathology, and, in some cases, genetics. The picture is incomplete but is already complex. PPA provides a good example of these studies. The PPA syndromes are generally associated with different cellular pathologies—nonfluent PPA mostly with tauopathies (abnormal intracellular inclusions containing a particular protein, tau) and fluent cases with ubiquitin inclusions (inclusions containing a different protein, ubiquitin). However, these associations account for only about half of the autopsied cases in each of the PPA syndromes; the remaining 50% of cases contain widely different pathologies, including pathology most commonly seen in other PPA syndromes (approximately 13% of fluent PPA cases had tauopathy in one study) and pathology associated with other diseases (ubiquitin inclusions in regions usually associated with motor neuron disease; plaque-and-tangle disease indistinguishable from that seen in Alzheimer’s disease, found in about 30% of cases in every PPA type). These results indicate that there are likely to be multiple etiologies of each of the currently recognized PPA syndromes. At the genetic level, siblings with an identical progranulin mutation diagnosed with PPA have varied in their clinical symptoms and course. Studies such as these show that an integrated psychological-neural-genetic approach to the neural basis for aphasia and to brain organization for language is complex.

Aphasia is now understood as consisting of deficits affecting linguistic representations and psycholinguistic processors. The relation of these deficits to brain areas, brain histology, genetics, and type of pathology is a topic of ongoing study.

David Caplan

Further Readings

Apraxia

Apraxia is a higher level motor disorder that patients exhibit when they perform voluntary movements. In right-handed individuals, lesions of the left parietal cortex produce apraxia of both limbs and buccofacial musculature (which includes the muscles of pharynx, larynx, jaw, velum, tongue, and lips). Less frequently, lesions of the premotor cortex and subcortical structures may also give rise to apraxia. There are a few cases of right-handed patients with apraxia following right-brain damage; however, the role of this hemisphere in higher motor control is not yet clear. In left-handers, apraxia is usually caused by lesions of the right hemisphere, may or may not be associated with aphasia—depending on

Author’s Note: This work was supported in part by a grant from NIDCD (DC00942).

See also Conduction Aphasia; Semantic Dementia
Apraxia

whether language in these individuals has migrated to this hemisphere—and occurs more rarely following lesions in the left hemisphere.

Apraxia is often but not always associated with aphasia—an acquired language disorder—as both action and language systems are predominantly lateralized in the left hemisphere. The fact that patients may have apraxia without aphasia, and vice versa, reinforced the view that these two deficits are not due to damage to a single mechanism. The anatomical contiguity of the cerebral regions in the left hemisphere that underlie action and language explains why lesions of these regions often impair both functions.

Types of Apraxia

According to Functions

The investigation of apraxia began over a century ago. The first to study this disorder systematically was Hugo Liepmann, who developed a model of how voluntary gestures are organized in the brain and how the ability to evoke and produce them can be affected by brain damage. Liepmann distinguished three different types of apraxia, according to the function that he hypothesized to be affected: ideational, ideomotor, and melokinetic. This clinical classification is based on the type of function or process that is believed to be damaged. Thus, ideational apraxia (IA) corresponds to a deficit in formulating an action plan (or movement formula); it manifests itself when patients are engaged in complex action sequences (as in preparing breakfast) but also when they use one object at a time (such as drinking from a glass). The ideational deficit has been attributed to different causes. Arnold Pick described patients with dementia who suffered from IA, making gross mistakes in using objects that they were able to identify. Subsequently, Joseph Morlaas suggested that IA was due to a faulty recognition of objects (i.e., agnosia of usage), and Klaus Poeck believed it reflected the disintegration of the sequence organization of object-related actions. According to this latter interpretation, patients should be impaired in all tasks tapping the representation of an action sequence (e.g., everyday activities and sequencing photographs of those activities), sparing the use of single objects.

Poeck's interpretation was challenged by the work of Ennio De Renzi and collaborators who reported patients who made mistakes, such as omissions, object misuse, and mislocations of objects, when they performed complex everyday activities as well as when they used individual objects. De Renzi interpreted IA as a conceptual disorder, caused by “amnesia” for object use that prevents patients from adequately accessing the functional characteristics of objects. However, the impaired ability to use objects is not necessarily caused by damage to semantic memory, as patients with tool use impairment and preserved knowledge about objects have been documented. Tool use deficits are prevalently associated with lesions of the posterior regions of the left hemisphere and, in particular, of the parietal cortex, a region that has been identified in imaging studies with healthy individuals as involved in tool use. IA is often caused by focal lesions of vascular etiology, but it can also be observed in patients with dementia.

After having correctly activated the action plan, a patient may fail to put it into practice. In Liepmann’s scheme, this deficit corresponds to what is generally known as ideomotor apraxia (IMA), but he originally called it motor and, later, ideo-kinetic apraxia. The best way to detect the presence of this deficit is to ask the patient to imitate the gestures performed by the examiner. Patients’ gestures may be inaccurate in different ways; instead of performing the desired gesture, patients may end up producing one that is a repetition of a previously performed gesture (perseveration) or substituting for the intended gesture one that does not bear any resemblance to it. Movement velocity, acceleration of the movement trajectory, and hand configuration may or may not be abnormal in such apraxic patients. These patients, in addition to the imitation deficit, may also have difficulties in generating gestures on verbal command. IMA is often produced by lesions of the left parietal and lateral premotor cortices, and of the corpus callosum (in this latter case, the deficit involves only the non-dominant hand). The ideomotor deficit is more frequent after lesions of the parietal cortex than of the premotor cortex and is, usually, more severe in the former case than in the latter. It has also been associated with lesions of the thalamus and basal ganglia and, more rarely, of the supplementary motor area.

The clinical distinction between IA and IMA has not always been accepted; some researchers favored
an alternate interpretation of IA as a more severe form of IMA. This interpretation, however, can be refuted on the grounds that the ideational and ideomotor deficits have been observed in a double dissociation: both patients with selective IA but without IMA and patients with the opposite pattern (i.e., with IMA but without IA) have been described. Double dissociations suggest that the abilities that are damaged in IMA and IA apraxia are at least partially different.

Liepmann mentioned a third type of apraxia, called melokinetic (or MKA, from the Greek melo, meaning “limb”), whereby patients perform awkward movements of fingers and hands contralateral to the lesion, regardless of the hemisphere involved, with apparently normal strength and sensation. This is a rare form of apraxia, prevalently produced by vascular lesions of the primary sensorimotor areas, although it has recently been observed in patients with corticobasal degeneration and Pick's disease. The existence of MKA has been put in doubt, as similar symptoms have been observed also in patients with mild paresis—a partial loss of movement—caused by corticospinal lesions.

**According to Body Effectors**

Apraxic disorders can be classified depending on the body parts whose movements are affected. Most frequently, patients are assessed for their ability to produce movements of the mouth and the upper limbs (either to verbal command or on imitation); sometimes movements of lower limbs and of the trunk are also included in the assessment.

Right-handed patients with lesions of the left hemisphere are normally found to have difficulties in performing organization of intentional movements of both upper limbs and, when tested, also of the lower limbs. Deficits at imitating finger configurations have been found in relation to anterior lesions of the left hemisphere, whereas imitation of hand postures can be affected by more posterior lesions. In contrast, it has been proposed that, whereas the left hemisphere is involved in imitation of movements of both the hand-arm and fingers, the right hemisphere seems to play some role in imitating only finger configurations.

Patients may also have difficulties in protruding the tongue, in whistling, and in performing other movements involving lips, mouth, and tongue, either on imitation or on command. This disorder is known as buccofacial apraxia (BFA) and occurs in the absence of paresis. Due to the contiguity of their neural underpinnings, BFA is often associated with apraxia of speech (or dysarthria, a deficit in programming the articulation of words); however, these two functions are somewhat independently organized, as double dissociations have been documented. BFA is produced by lesions of the premotor cortex and of the anterior part of the insula in the left hemisphere. Finally, trunk apraxia produces deficits in performing axial movements (involving the longitudinal body axis) that are often associated with limb apraxia and gait apraxia; it is observed following bilateral frontal lesions.

**Other Types of Apraxia**

Callosal apraxia has both an ideational and an ideomotor component, and affects the movements of the limbs ipsilateral to the dominant hemisphere, while those of the contralateral limbs are spared. Patients with nonsurgical lesions of the corpus callosum may show apraxic deficits, particularly if the fibers that connect the supplementary motor areas, the lateral premotor areas, the superior and inferior parietal lobes, and the primary sensorimotor areas are affected. In contrast, movements of patients with agenesis of the corpus callosum (a congenital disorder in which there is a complete or partial absence of the corpus callosum) are surprisingly accurate, except for those that require a bimanual coordination, while those of patients who have undergone commissurotomy (whereby the neurosurgeon severs communication between brain hemispheres to prevent the transfer of seizures in epileptic patients) appear awkward only when they are performed by the left hand on verbal command or on imitation of a stimulus presented in the right hemifield. This relative preservation might be due to some degree of reorganization of the nervous system.

**Constructional apraxia**, a term coined by Karl Kleist (1879–1960), comprises a heterogeneous set of deficits in constructing, composing, and designing complex structures such that the spatial form of the product is not adequate. Constructional apraxia is independent of limb apraxia and is caused by either left or right hemisphere lesions.

Dressing apraxia arises when one or more of the following forms of knowledge is affected by brain
Apraxia damage: how to use different clothes; in which context; the motor knowledge necessary to put them on; and the matching knowledge between body parts and clothes. It can also be caused by unilateral neglect—a neuropsychological condition in which, after damage to one hemisphere of the brain (more frequently the right hemisphere), the patient shows an inability to attend to the side of space or body contralateral to the lesion.

Models of Apraxia

Although Liepmann’s original schema has subsequently been modified, the essential dichotomy between the two levels of movement production is still largely used, especially for clinical purposes. In 1991, Leslie Rothi, Cynthia Ochipa, and Kenneth Heilman proposed a cognitive neuropsychological model of limb praxis similar in logic to models of language production. The main features of the model are depicted in Figure 1.

The model distinguishes between action inputs and action outputs, but it also comprises intermediate action processing stages. A visual gesture can be imitated using a semantic or a non-semantic route. After visual analysis, the semantic route encompasses the input praxicon, which, as it contains all the gestures we know, allows recognition of a gesture as belonging or not to one’s own repertoire, and the semantic or conceptual system, which stores the meaning of gestures, while the output praxicon permits one to actually produce them. Whereas the semantic route supports imitation of meaningful gestures, the non-semantic route is necessary for imitating novel gestures. The model also accounts for how we use objects and the way we pantomime their use, either in the visual or verbal-auditory modality.

Evidence from neuropsychological observations supports this cognitive architecture. Thus, any damage involving the input praxicon impairs the ability to discriminate between known and new gestures (pantomime agnosia), leaving intact the ability to perform semantic tasks and to produce gestures on imitation and on verbal command. Damage to the semantic system leads to impaired performance in semantic tasks (also referred to as ideational or conceptual apraxia) and to impaired pantomiming on verbal command. Damage to the output praxicon causes a deficit in all tasks involving the production of gestures, except for the imitation of novel gestures, which is affected only when the non-semantic route is damaged (also called visuo-imitative apraxia).

The main modifications subsequently made to the original model by other scholars are the following: the substitution of the innervatory patterns by a temporary memory system for gestures and the acknowledgment of an explicit role of a body representation, although its exact nature is not clear.

The idea that limb apraxia occurs only when the gestures are requested by the examiner but disappears when the same gestures are triggered by the
Argument Mapping

Argument mapping is diagramming the structure of argument, construed broadly to include any kind of argumentative activity, such as reasoning, inferences, debates, and cases. This entry briefly surveys the nature, benefits, and historical context of this activity.

Nature of Argument Mapping

An argument map is typically a “box and arrow” diagram with boxes corresponding to propositions and arrows corresponding to relationships such as evidential support. Argument mapping is similar to other mapping activities, such as mind mapping and concept mapping, but focuses on the logical, evidential, or inferential relationships among propositions. Argument mapping is concerned with informal reasoning and “real world” argumentation and thus contrasts with the use of diagrammatic techniques in formal logic such as Venn diagrams.

Argument mapping is done in many different ways. Any such approach is usefully seen as making commitments at three different levels: theory, visual conventions, and technology.

First, argumentation theory and related fields such as informal logic, critical thinking, and rhetoric provide the theoretical framework for any style of argument mapping. The theory specifies the entities, relationships, and values to be represented and provides rules or guidelines governing map construction. Conversely, developing an argument mapping scheme can force theoretical issues into the open and stimulate further research.

Second, an approach must adopt visual conventions for displaying arguments in accordance with the theory. From the range of dimensions such as shape, color, and line, the scheme designer must choose, for example, how to show that one proposition supports another. The conventions should yield maps that not only are theoretically adequate but also communicate effectively, properly conveying to the reader the argument structure and associated issues; support interaction (construction and modification); and please the eye. It is challenging to satisfy all of these constraints, and a good scheme will draw on fields such as information visualization and cognitive science.

Third, creating argument maps requires resources and technology of some kind. The most obvious and accessible technologies are pen and paper or whiteboards, but these quickly reveal their limitations, which include poor support for complex diagrams and modification of diagrams, and failure to constrain, scaffold, or guide the user in any way. Serious argument mapping is now done using specially designed computer tools that have emerged over the past decade, such as Araucaria, Compendium, bCisive, and SEAS. Such tools largely overcome the disadvantages of manual technologies and offer a number of other advantages. An important recent development is the growing array of collaborative online argumentation tools such as Debategraph,
although good user interfaces for online argument mapping remain an important challenge.

Benefits of Argument Mapping

Argument mapping enthusiasts believe that argument mapping has numerous potential benefits. For the suitably skilled person, mapping a complex argument promotes clarity and insight, more rigorous and complete articulation, and more judicious evaluation. Teachers use argument mapping to help students acquire basic concepts, better understand how arguments are constructed, and enhance their reasoning skills. Argument mapping can be an effective way to improve general critical thinking skills. In the workplace, argument mapping can promote rational resolution in complex, fractious debates; improved communication of important arguments; and better decision making.

There is a simple, plausible case as to why argument mapping should have these benefits. Arguments can be complicated, sometimes extremely so. We know that well-designed visualizations help our minds cope with complexity. Hence, argument maps should help us deal with arguments. This can be easily demonstrated with simple exercises, but there is scope and need for rigorous empirical investigation.

History and Future of Argument Mapping

It appears that argument mapping did not emerge until the 19th century, with the first reference to the practice in a logic textbook by Richard Whately in 1836. The most notable early exponent was the legal theorist John Henry Wigmore, who developed detailed schemes for displaying legal evidence in the early 20th century. In the mid-20th century, philosopher Stephen Toulmin developed a simple but influential argument diagramming scheme. With the rise of the informal logic and critical thinking movements, argument mapping began to make regular appearances in textbooks. Interest in argument mapping grew rapidly in the 1990s, due in large part to the increasing availability of computers and specially designed software. A substantial series of maps released by Robert Horn stimulated widespread interest in the technique. In recent years, argument mapping has crossed over from academic or educational applications into the workplace (e.g., in intelligence analysis and policy development) and popular use, particularly with the recent emergence of systems for online collaborative mapping.

The recent surge in argument mapping signposts an exciting development in humans’ cognitive history. Computer-supported argument mapping offers
a major change in the way we handle informal reasoning and argumentation. It is an instance of Douglas Engelbart’s vision of the augmentation of human intellect, whereby we develop technologies that can boost our individual and collective intelligence by complementing our own cognitive machinery. Contemporary argument mapping approaches and tools are still relatively rudimentary. We can look forward to more sophisticated frameworks integrating more seamlessly with our biologically endowed cognitive equipment, enhancing our capacity to deal with future intellectual challenges.

Tim van Gelder

See also Distributed Cognition; Extended Mind; Thinking

Further Readings


Atomism About Concepts

Concept atomism is the view that most lexical concepts (i.e., most concepts associated with single words in natural languages such as English) are not composed from other concepts. They are semantically unstructured, or primitive. For example, the concept dog is not built from furry, quadruped, animal, or any other concepts. Consequently, a thinker can possess the concept dog without possessing any other concepts in particular, and can thereby think that Lassie is a dog without being able to think that Lassie is furry, that Lassie is a quadruped, that Lassie is an animal, and so forth.

Concept atomism was first formulated and defended by philosopher Jerry Fodor. On his version of the view, concepts are symbols in a language of thought that acquire their referents by virtue of standing in appropriate causal relations to those referents. For example, the concept dog is a mental symbol that refers to dogs not because of its relation to other concepts, but because of its causal relation to dogs.

Motivations for Atomism

One motivation for concept atomism is the suspicion that, dictionaries notwithstanding, most lexical concepts cannot be rigorously defined. Consider, for example, the concept justice. Ever since Socrates, philosophers have struggled to define this concept, and with little apparent success. Yet if justice were built from other concepts, we should be able to say what those concepts are. We should be able to define it. Atomists thus conclude that justice is a primitive concept.

Some philosophers have replied that justice is anomalous and have pointed to concepts that seem to be easier to define. For example, the concept bachelor might seem to be definable in terms of the concepts unmarried and man. But atomists are skeptical. Imagine a 60-year-old grandfather who never took his vows but has been in a committed, monogamous relationship with the mother of his children for 40 years. It seems inappropriate to classify him as a bachelor. Or consider the pope, who is certainly an unmarried man but, again, does not qualify as a bachelor. Atomists argue that examples such as these show that even concepts such as bachelor are ultimately indefinable and thus primitive.

A second motivation for atomism derives from the apparent fact that people can think about things about which they are ignorant or misinformed. For example, philosopher Hilary Putnam claims to be ignorant of the difference between elm and beech trees. But when he thinks, “This forest contains elms,” he is nevertheless thinking about elms, not beeches. Atomists take this to show that the concept elm cannot be structured from other concepts (e.g., tree with doubly serrate leaves) because, when Putnam thinks about elm trees, his ignorance prevents him from employing those other concepts.

Objections to Atomism

Opponents of atomism are not persuaded by these considerations. They argue that, although concepts such as bachelor and dog may be difficult to define, it doesn’t follow that they are indefinable; they argue
also that any person who doesn’t know the difference between an elm and a beech doesn’t really have a concept of either. Additionally, opponents of atomism point to two counterintuitive consequences of the view.

First, it is natural to assume that learning a concept is a matter of learning which concepts it is built from. For example, a child acquires the concept bachelor by learning that something is a bachelor if and only if it is unmarried and a man. But if concepts are unstructured, this picture of concept learning is in trouble. For if the concept bachelor isn’t composed from the concepts unmarried and man, then acquiring the concept bachelor cannot be a matter of learning that bachelors are unmarried men. Concept atomism thus conflicts with a tempting view of concept learning.

Not all atomists are troubled by this conclusion, however. On the one hand, Fodor embraces the view that most lexical concepts are unlearned. This does not necessarily mean that we are born with concepts such as bachelor and carburetor, but it does lump them together with sunburns, pubic hair, headaches, and other things that we acquire in life without learning. On the other hand, atomists such as Eric Margolis and Stephen Laurence attempt to defend the idea that concepts are learned by specifying methods of concept learning that do not involve uncovering a concept’s constituent structure.

A second worry about atomism concerns its ability to distinguish coextensive concepts, such as water and \( H_2O \), Hesperus and Phosphorus, or Mark Felt and Deep Throat. Given that the identification of each of these pairs was newsworthy, it seems reasonable to count each concept as distinct. But given that the members of each pair refer to the very same thing, it is less clear what makes them distinct. On a fairly standard account, these various concepts are distinct because they have different components. For example, the concept water might be composed from such concepts as clear, drinkable, and liquid, and the concept \( H_2O \) from hydrogen and oxygen. Notice, however, that this standard account of what distinguishes coextensive lexical concepts will usually be unavailable to the concept atomist because the atomist holds that most lexical concepts are not built from other concepts. The atomist who doesn’t want to deny the seemingly obvious truth that these various concepts are distinct must therefore find some other way to distinguish them. One possibility, embraced by Fodor, is to appeal to their differences at the symbolic level. For example, what makes the concepts Deep Throat and Mark Felt distinct is their association with different symbols in the brain—for example, different tokens in a language of thought. But the plausibility of this approach, along with concept atomism more generally, remains controversial.

*Jacob Scott Beck*

See also Causal Theories of Intentionality; Classical Theory of Concepts; Concepts, Philosophical Issues

Further Readings


attention that reflect the unique adaptive problems of
a species. For example, in humans, attention is biased
toward faces over other sorts of images. Recent work
by Alan Bond and colleagues suggests that physical
characteristics of a prey species may be shaped by the
attentional limits of a predator. Blue jays may hunt
using a search template for a locally frequent form of
moth, producing selective predation (and subsequent
disadvantage) for the frequent over the rare form.
The consequences of attentional limits, therefore, can
be widespread, and the morphology of a species is
shaped in important ways by the attention systems of
its conspecifics and its predators.

The Ubiquity of Selective Attention
In humans, the neuroscience of selective atten-
tion emphasizes the contributions of large cortical
networks. It is clear that an intact cortex is neces-
sary for efficient selective attention in humans, and
diverse forms of cortical damage produce many
kinds of selective attention disorders, such as hemi-
spatial neglect, in which patients may not respond
to objects in the left side of space (e.g., they may
refuse to eat food from the left side of the plate or
fail to read words on the left side of the page). Many
brain imaging studies show that even apparently
simple tasks, like shifting attention from one object
to another, activate a large network of cortical areas.

However, the cerebral cortex is a very special
structure in evolutionary terms. Within the mam-
mals, the cortex is significantly larger in primates
than non-primates, and larger still in humans than
other primates. Selective attention in humans, there-
fore, requires machinery that is, in evolutionary
terms, unique and expensive. On this basis, we might
expect selective attention to be crude or even absent
in other species lacking a well-developed cortex.
However, this intuition would be mistaken: Selective
attention is far from uniquely human and appears to
operate to some degree in even the smallest brains.

Selective attention abilities and costs of various
kinds have been documented in many diverse
nervous systems, as indicated by the examples in
Table 1. Inclusion in the table is based on whether
the brain and overt response to a stimulus have been
shown to depend on the animal’s current state and
behavioral goals. This is a broad definition, and the
table entries represented are not perfectly compa-
rable. For example, fruit flies are included in the
table because they respond more selectively to visual
motion signals when they are engaged in flight.
Frogs selectively respond to one prey stimulus when
there are multiple prey present. Selective processing
by blue jays has been documented in a variety of
tasks, including costs for switching between targets,
and the use of templates in visual search for prey.
The distance between human brains and those of
other species can be expressed both in time (millions
of years since the last common ancestor), and space
(ratio of brain volumes). There are always different
concerns about when it is appropriate to adjust for
body size when comparing brain sizes. Table 1 pre-
sents brain sizes unadjusted for body size. The rea-
son is to better indicate just how small some of the
brains are that can be engaged in selective attention.
What is clear, even from this incomplete sketch, is
that attention systems have been in place a long time,
they are prevalent throughout the animals, and they
can operate, at least to some degree, in the absence
of large and evolutionarily expensive machinery like
the human cortex.

Even simulated organisms, evolved through unsu-
pervised learning, are capable of attending selectively
to their environment. For example, Robert Ward
and Ronnie Ward used genetic algorithms to create
an artificial agent, with a “brain” of only eight sim-
ple units, which could perform some sophisticated
selective attention tasks. A number of agents with
randomly wired “brains” were assessed on a series
of selective attention problems involving two target
objects that had to be “caught.” The agent had a
simple sensorimotor system of proximity sensors
and locomotion. The agents that performed best
on the catching tasks were recombined to produce
a new generation, which itself was assessed for the
best performers, and so on. The result of this process
over many generations was a simple artificial agent
that could inhibit its responses toward distractor
objects, and then subsequently release that inhibi-
tion so that a former distractor could guide behav-
ior. In so doing, the agent also created a memory
for unselected items, and demonstrated the counter-
intuitive effects of reactive inhibition, in which it
takes longer to orient to previous distractors when
they are highly salient.

The ubiquity of selective attention systems,
coupled with the apparent neural complexity of this
system in humans, raises some puzzling questions.
Why are such large networks in the human system
necessary? In what ways do people differ from other animals in their selective attention requirements? The human literature offers little guidance here, because an interesting but largely ignored question is whether attention phenomena found in humans address the properties of selective attention systems in general, or of human attention specifically. However, many of the standard attention operations which have been most scrutinized in humans—searching for and prioritizing targets, suppressing the influence of distractors, shifting attention from one object or location to another—seem just as important for a frog, fish, or fly, as for a human. For example, consider an animal placed between two potential prey. Selective processing allows the animal to direct its behavior toward one of the targets and ignore a tendency to respond to the other. A failure of selection could mean a response directed toward some intermediate but useless point between the two. Such a nonselective animal would be cruelly exposed for sensory exploitation from its competitors.

One way to reconcile the neural complexity of human attention with the ubiquity of selective attention in even very small brains is to suggest that attention reflects a state of the nervous system rather than a specific process. To illustrate, consider again the large activations found in parietal and frontal cortex when people shift their attention from one object to another. It is tempting to think of the activated network as something like the gears and cogs needed to move the focus of the tower searchlight from one corner of the prison yard to the next. But why would millions of neurons be necessary for this fundamental operation when it can be readily achieved in species with much more limited nervous systems? It seems more sensible to work from the hypothesis that such large activations reflect the consequences of the attention shift, a conclusion reached through other means in John Duncan’s theory of biased competition. There is no specific “attention” module in a biased competition model; instead, the activity of multiple brain regions, specialized for different tasks, are coordinated by their interconnections to focus on a single object and its implications for behavior.

### Has Attention Evolved Repeatedly?

Given the range of species with at least some ability for selective attention, a natural question is whether attention circuits in different species reflect a highly conserved, or shared, neural system. Alternatively, would it be reasonable to interpret the diversity of systems as evidence that attention evolved repeatedly in the history of life? This approach is certainly plausible, and some data do support it. However, I believe it would be a mistake to think of attention systems as evolved independently in different lineages. Attention systems are not unique to a single species, and they are not the exclusive domain of vertebrates. As an evolutionary perspective, attention systems are best understood as a necessary part of the nervous system in every organism having the capacity for behavior. The neural complexity of attention systems varies among species, but attention is a fundamental aspect of behavior across the phylogenetic range.}

---

**Table I** A selection of nervous systems, tested and demonstrated to be capable of various acts of selective attention, compared to the human brain in terms of approximate time of divergence from the human evolutionary line and relative size of brain

<table>
<thead>
<tr>
<th>Species</th>
<th>Approximate Time of Divergence From Humans</th>
<th>Relative Brain Size (Unadjusted for Body Size)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humans</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Chimpanzee (ape)</td>
<td>6 million years ago</td>
<td>33%</td>
</tr>
<tr>
<td>Macaque (monkey)</td>
<td>25 million years ago</td>
<td>7%</td>
</tr>
<tr>
<td>Rat (non-primate mammal)</td>
<td>75 million years ago</td>
<td>0.5%</td>
</tr>
<tr>
<td>Blue jay (bird)</td>
<td>310 million years ago</td>
<td>0.5%</td>
</tr>
<tr>
<td>Frog (amphibian)</td>
<td>340 million years ago</td>
<td>0.01%</td>
</tr>
<tr>
<td>Fruit fly (invertebrate)</td>
<td>500 million years ago</td>
<td>0.0001%</td>
</tr>
<tr>
<td>Artificial agents</td>
<td>N/A</td>
<td>0.00000001%</td>
</tr>
</tbody>
</table>


Note: Brain sizes here are not adjusted for body size, to illustrate how brains that are very small relative to a human can still perform acts of selective attention. The correlation between divergence and brain size in the tested species is merely accidental. For example, a whale is a non-primate mammal with a larger unadjusted brain size than a human.
perhaps the circuitry for attention, while performing analogous functions in different species, has evolved independently, in much the same way that analogous solutions have evolved for other “good ideas,” like flying. A definitive answer awaits a comparative approach, which considers similarities and differences in the neural organizations and behaviors of species, as well as their genetic makeups. However, it is clear that if selective attention has evolved repeatedly, we should expect a variety of neural mechanisms adapted for an animal’s specific circumstances. Such a diversity of circuitry has already been demonstrated in a domain at least related to attention. Animals benefit by knowing to what extent the changing stimulation across their sensors is due to their self-motion, and to what extent this pattern reflects motion within the environment. By integrating sensory signals with information about current motor commands, referred to as corollary discharge, it is possible to separate the effects of self and environmental motion. This solution allows an animal to ignore the sensory influence of self-motion, while selectively processing the changes in the environment. A recent review by Trinity Crapse and Marc Sommer illustrates a wide range of corollary discharge circuitry across species, each designed to work with the content and consequences of a species-specific behavioral repertoire. The diverse mechanisms for processing corollary discharge illustrate how a problem that is universal for any nervous system (separating self from environmental stimulation) is subject to highly specific solutions tied to the unique circumstances of individual species. That is, although the problems relating to corollary discharge are universal, it seems plausible that solutions to the multiple different solutions to the problem have evolved.

The literature on human attention is rich with phenomena, but for the most part, comparative studies on these phenomena, such as the attentional blink or psychological refractory period, have yet to be done. An exception is work with artificial agents. In some ways, the comparison between human and artificial systems is very powerful, as any similarity in attention cannot be based on shared inheritance but solely on the computational realities of selective processing. On the one hand is an artificial system of a few simple processing units, evolved to operate in an environment consisting of only two other objects. At the other extreme is the human system, with many billions of neurons, adapted for an extremely complex environment. If both systems use a common computational approach, this is unlikely to be a coincidence. Instead, the common mechanism may be a fundamental computational mechanism for selective processing, in much the same way that lateral inhibition is a mechanism for contrast enhancement in many neural systems. In any case, it is clear that comparative approaches can produce one of two results, and both are valuable. Either the human and nonhuman systems work in similar ways, in which case the nonhuman system validates and extends the human literature, or the nonhuman system operates according to new principles not yet seen in the human literature. In this case, we have a working model for new experimentation and theory.

Robert Ward

See also Attention, Neuroimaging Studies of; Natural Action Selection, Modeling

Further Readings
as electroencephalography (EEG) and magnetoencephalography (MEG), and hemodynamic methods, such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI). Electromagnetic methods measure the voltage or magnetic field changes that are produced by neural activity, through electrodes attached to the scalp (in EEG) or sensors located close to the scalp (in MEG). These methods provide excellent temporal resolution, recording neural activity in real time, but have limited spatial resolution because the measured response is recorded at some distance from the neural activity and each sensor may measure activity from multiple brain regions simultaneously. Hemodynamic measures (e.g., PET and fMRI) provide complementary information, with high spatial resolution but limited temporal resolution. This is because these methods rely on blood-flow-related changes in the brain that lag behind the actual neural activity by several seconds. For PET, variations in blood flow can be measured after injection of a radioactive isotope into the blood, whereas fMRI relies on intrinsic magnetic properties of blood. Although PET and fMRI provide excellent spatial resolution, the temporal resolution is limited and the precise timing of the hemodynamic response can vary across regions, further complicating the ability to understand the dynamics between active brain regions. As reviewed in this entry, electromagnetic and hemodynamic methods have each provided a wealth of information regarding the neural architecture of attention systems, and combining these methods can provide even more precise information on the spatiotemporal dynamics of attention mechanisms in the brain. The present entry primarily focuses on the extensive neuroimaging work on visual attention. It is divided into two main sections, reviewing work on the effects of attention (i.e., the “sites” at which the effects of attention are manifest) versus the control of attention (i.e., the “sources” of attentional control).

The Effects of Attention

“Sites” of Attentional Modulation

A fundamental issue in attention research is the stage of processing in the brain at which attention can begin to affect processing. The results from numerous neuroimaging studies converge on the general conclusion that there is considerable flexibility in the site of attention effects, although there may be a limit to how early these effects can occur. These studies have revealed that attention cannot simply be characterized as a unitary mechanism that consistently acts at a single, fixed site of action. Rather, a variety of sites, both early and late in the processing stream, have been found to be affected by attention, and the precise site of attention at any time is a function of multiple factors. For example, paying attention to the spatial location of a stimulus has been shown to affect early stages of cortical processing, within the extrastriate fusiform and middle occipital gyri. Attending to nonspatial features (e.g., color, motion, faces) modulates neural activity most strongly in the later, higher order areas that specialize in processing those features. These imaging results highlight that attention is flexible and that the sites of attention effects vary depending on what’s being attended.

PET and fMRI methods alone have not provided a definitive answer to the question of the earliest point in the brain at which attention can affect processing. Although numerous studies have found robust effects of spatial attention in extrastriate areas, a few neuroimaging studies have also found attention effects in striate cortex. Before concluding that attention can modulate visual processing as soon as it first enters the cortex, however, it is critical to consider the temporal imprecision of PET and fMRI measures. These methods rely on measuring the relatively sluggish hemodynamic response that lags behind neural activity by multiple seconds. Furthermore, visual processing is not simply feedforward through successive visual areas but also includes subsequent feedback to earlier processing areas from higher order areas. Therefore, the lack of temporal precision limits the conclusions that can be drawn from PET and fMRI. For example, an attention effect in striate cortex could reflect the initial first-pass processing of sensory information into the cortex, or it may be due to effects occurring well after that initial processing. Indeed, decades of research using event-related potentials (ERPs; stimulus-locked averaged sections of the electroencephalogram) have provided evidence that the earliest effect of attention occurs after the initial processing through striate cortex has proceeded without modulation. To address these apparently conflicting results regarding the earliest effects of attention, it is helpful to consider results from studies that have combined fMRI and ERP measures.

As described earlier, ERPs have complementary strengths and limitations compared to PET and
Attention, Neuroimaging Studies of fMRI. Therefore, combining these methods can provide both high spatial and high temporal precision. The ERP component that reflects the initial processing in striate cortex is labeled the C1 ("C" for its central scalp location), whereas the first major ERP index of extrastriate processing is a subsequent component labeled the P1 ("P" referring to its positive polarity). Critically, ERP studies have consistently found the earliest attention effect to be a modulation of the P1 component, whereas the earlier C1 component has not typically been found to be modulated by attention. However, some fMRI studies have found attention effects in striate cortex. In a landmark study combining ERPs with neuroimaging, Antigone Martinez, Steve Hillyard, and their colleagues found that the earliest effect of spatial attention occurred at approximately 75 milliseconds latency (associated with the P1) and was localized to the extrastriate fusiform gyrus, whereas fMRI-measured attention effects in striate cortex were found to be associated with ERP effects occurring significantly later in time, well after the C1 and P1 components. The initial processing in striate cortex that gave rise to the C1 component was unaffected by attention. These data provide evidence that attention effects in striate cortex, at least in some fMRI studies, do not reflect the initial feed-forward activity, but rather reflect later re-entrant processing in that region. Recent ERP studies have found attentional modulation of the C1 component, although those studies did not use fMRI or PET to definitely locate the neural site of these effects. Further studies are required to fully understand all the parameters that determine the site of attention in any given situation. However, neuroimaging studies have shown that the site of attention effects is flexible and reflective of perceptual and task demands. The following sections review work on the attentional control network in the brain that is thought to be responsible for producing the attention effects described earlier.

The Control of Attention

Orienting Attention ("Sources" of Control)

Mechanisms of attentional control are thought to precede and to be responsible for the effects of attention described earlier. Neuroimaging studies have revealed that a widespread network underlies attentional control, with two key areas being the frontal eye fields (FEF) and the intraparietal sulcus (IPS). This frontal-parietal network has consistently been found to be active in studies of voluntary attention across a variety of conditions, including overt attention as well as covert attention, spatial attention as well as object attention, and visual attention as well as auditory attention. Although numerous factors determine the precise site of attentional effects, the control of attention has been shown to consistently engage this frontal-parietal network.

A prominent model of attention holds that there is right hemisphere dominance for attentional control. Neuropsychological patient studies have generally found that damage to the right hemisphere is often more debilitating on attention tasks than is damage to the left hemisphere. To explain this pattern of results, it has been postulated that, in the healthy brain, the right hemisphere monitors both visual fields equally well, whereas the left hemisphere monitors the contralateral right visual field more strongly than the left visual field. Most fMRI studies of attentional control, however, have found the frontal-parietal network to be engaged equally strongly bilaterally for both visual fields. A recent fMRI study by Joseph Hopfinger, Christine Camblin, and Emily Parks, however, revealed a hemispheric specialization in healthy participants that is in accord with the asymmetry seen in patient studies. In that study, voluntary attention was isolated by measuring self-initiated shifts of attention, in contrast to previous studies that used an abruptly appearing cue stimulus to instruct subjects where to attend. For self-initiated shifts of attention, the frontal-parietal network in the right hemisphere was equally active for attending to both the left and right visual fields, whereas the left hemisphere control network areas were much more strongly active when attention was oriented to the contralateral right visual field compared to the ipsilateral left visual field. Thus, this neuroimaging evidence supports theories of hemispheric asymmetry of attentional control based on patient studies and extends it to the healthy adult brain.

Although neuroimaging studies have consistently found a frontal-parietal network involved in attentional control, the sequence of activity between these regions is difficult to resolve with fMRI alone. Early ERP studies suggested a parietal-to-frontal progression, but single-unit recording studies in nonhuman primates have suggested a frontal-to-parietal progression, at least for voluntary attention. In an
important study combining ERPs and fMRI, Tineke Grent-’t-Jong and Marty Woldorff provided new evidence from healthy adults that frontal regions appear to initiate the process of attentional orienting, with parietal regions coming online later to instantiate the shift of attention.

Beyond the localization of the attentional control regions, another question is how this frontal-parietal control network produces the attention effects described in the first part of this entry. fMRI studies by Sabine Kastner and colleagues, and by Joseph Hopfinger and colleagues, have shown that there are significant and robust effects in visual processing regions that precede the presentation of the target stimuli. These studies have shown that the same visual areas that show robust attentional effects during the processing of the target stimuli show similar attention effects before the target stimuli appear. Thus, a critical mechanism of attention appears to be the biasing of sensory processing regions before the appearance of stimuli, presumably preparing those areas of the brain so that incoming information will be selectively processed according to the goals of the observer. In summary, neuroimaging has provided evidence that voluntary attention relies on a frontal-parietal network that biases relatively early sensory processing areas in preparation for selectively enhanced processing of the attended regions and objects.

Disengaging and Reorienting Attention

The frontal-parietal attentional control network described in the previous subsection is most closely associated with the voluntary shifting of attention. This final subsection reviews the situation when voluntary control fails—when attention is involuntarily captured to the wrong location and must be reoriented back to the desired location. Following a reflexive capture of attention, one needs to disengage from the distracting stimulus and reorient attention back to its original focus. Recent work by Maurizio Corbetta and colleagues, among others, has provided evidence that the systems underlying disengagement and reorienting are distinct from the frontal-parietal network described earlier. Whereas voluntary attentional control involves dorsal regions of the frontal and parietal cortex, reorienting of attention evokes activity in more ventral regions of the frontal and parietal cortices; the temporal parietal junction, in particular, is strongly implicated in the reorienting process. Neuroimaging studies have not provided much evidence regarding the sequence of frontal and parietal activity in the disengagement and reorienting of attention; however, studies with transcranial magnetic stimulation and single-cell recordings suggest that reflexive attention involves a parietal-to-frontal sequence in contrast to the frontal-to-parietal progression associated with voluntary attention.

In summary, electromagnetic and hemodynamic neuroimaging methods have provided key insights into the neural mechanisms of attention. A distributed frontal-parietal network is involved in controlling the orienting and focusing of attention, leading to selective biasing of sensory processing regions before the expected targets appear. The exact site of attentional effects is flexible, depending on perceptual and task demands, but appears to be most robust within early extrastriate regions for visual spatial attention. Finally, regions of the frontal and parietal cortices ventral to those involved in the initial orienting of attention seem to be involved when attention needs to be reoriented following an involuntary capture of attention toward a distracting stimulus. Future studies using these methods, separately and in combination, should provide further insight into the dynamic brain mechanisms that allow us to voluntarily focus our attention and reduce potential distraction.

Joseph Hopfinger

See also Attention and Action; Electrophysiological Studies of Mind; Neurodynamics of Visual Search

Further Readings


**ATTENTION, RESOURCE MODELS**

This entry addresses models of processing resources developed by researchers who study attention, largely to account for limits of performance. A prominent performance limit is the common difficulty of attending concurrently to more than one object or of conducting more than one cognitive process at the same time, typically encountered in multiple-task situations (namely, in which a person is to perform more than one task at a time).

Whereas attention is clearly associated with selection, namely, the operation that enables the execution of one particular controlled process (rather than others) on one particular object (rather than on others), there is no agreement on what requires selection or on what limits it.

One stance, asserted in what Diego Fernandez-Duque and Mark Johnson call “cause theories,” holds that attention is enabled by some internal input required for processing (e.g., locations in working memory, share of dedicated time in the executive control system, communication channels). The allocation of that input determines both what is selected and the quality of behavioral output of whatever is selected. The prevalent postulate of such theories is that performance, or the information processing that must generate it, is enabled by the availability of that scarce internal input, and hence performance deficits are due to the fact that it is not supplied in sufficient quantity, or sufficiently early, to the object pertinent for that performance. The volume of that internal input is often called processing resources or cognitive resources.

**Evolution of Models**

The notion of resources explicates a premise implicit in naive thinking and reflected in natural language: Attention is said to be “paid,” namely, directed and/or expended. Whatever is expended must be taken to be some sort of structure or commodity that is limited in some way, hence internal input. Early theorists of attention, such as Alan Welford and Donald Broadbent, construed input, respectively, as a uni-partite channel that may only be directed as a whole to a task or an object of attention, or as a single channel with limited capacity that is typically capable of handling only a single task.

Daniel Kahneman posited that internal input may be expended in various amounts and may be divided between objects of attention. He argued that the existence of such input is suggested by what he called the intensive aspects of attention, namely, the dependence of performance on vigilance or arousal, which he ascribed to the degree of accessibility or activation of that internal input.

Donald Norman and Daniel Bobrow coined the term resources as a label for the hypothetical internal input assumed to be demanded for a broad range of functions believed to require attention. They suggested that performance of a task or process is resource limited, namely, that it depends on the amount of resources invested in it, up to a point at which performance becomes data limited, namely, insensitive to resources and determined just by data quality. For example, a task of detecting a moderately loud tone may demand a small quantity of resources; when that quantity is not fully available, detection may sometimes fail; when it is available, failures must be due to poor sensory data. However, the shape of this performance-resource function is strongly task specific. The way Norman and Bobrow proposed to study that function is via analyzing dual-task performance by means of a curve they termed performance operating characteristics (POCs), plotting performance of one task as a function of performance of the concurrent task.

In the first formal model of resources Norman and Bobrow proposed, attention was roughly conceived of in terms of allocation of a single type
of that hypothetical internal input demanded for processing that is available in a limited quantity. Subsequently, David Navon and Daniel Gopher formulated a comprehensive theory of processing resources couched in terms of analogies from microeconomics, meant to jointly encompass accounts of various effects on performance—those of incentive, task difficulty, task load, task priority, arousal and vigilance, skill, and so on. The theory considered two hypothetical cases: the case of a single, all-purpose resource pool and the case of multiple pools, elaborated to accommodate modality-specific (visual, auditory, etc.) and analysis-specific (encoding, response selection, etc.) capacities, any combination of which may be demanded for the performance of a specific task or process. Obviously, the particular combinations are determined by the nature of the tasks and hence may vary between tasks. Navon and Gopher pointed also to the intricate dependence of the interpretation of data presented as POCs on various assumptions about processing in dual-task situations. Christopher Wickens concluded from the examination of a number of dual-task studies that there are about a dozen sources of task interference that can be characterized as conjunctions of values of three variables: (a) stage of processing (perceptual, central, response), (b) code of processing (verbal, spatial), and (c) modality (of input—visual, auditory; of response—manual, vocal). He interpreted that as constituting evidence for the existence of at least the same number of separate pools of resources.

As explicated somewhat later by Navon, resource theory must presume that resources are defined by five properties:

1. Aggregate nature. Resources come in units, any number of which can be invested in a certain process.
2. Exclusive usage of units. Resources are private commodities in the sense that each unit can be used or consumed by only one process at a time.
3. Distributability. Resources can be invested in more than one process at a time, in that different units are allotted to different processes.
4. Effectiveness. The number of resources invested in a single process affects the quality of its output.
5. Scarcity. The number of resources available at any point in time is limited.

Although resource models had gained popularity in cognitive psychology and human factors research as of the late 1970s, they lost quite a bit of support after the mid-1980s, when critics argued that, despite their rigorous appearance, they bear predictions that are barely falsifiable, hardly selective, sometimes little beyond the obvious, and very difficult to test. Alan Allport first noted the circularity in the predictions of early resource theory. In 1984, Navon showed that almost all of the phenomena predicted from resource theory (effects of incentives, effects of task difficulty or complexity, dual-task deficits, effects of task priority, effects of concurrent task difficulty/complexity, and interaction of difficulty/complexity with manipulations meant to tap resource availability) can be accommodated by an alternative notion that does not resort to the five properties of resource theory. Instead, the latter notion postulates that (a) a process may be variably activated, and (b) the output quality of the process depends on the states, varying in time, of various cognitive or bodily variables called alterants (e.g., anxiety, activation of a certain representation in memory). Navon also worried that resource theory, despite its appeal due to its apparent rigor, would not prove sufficient to impose conceptual organization on phenomena in the field. In that case, the concept would contribute very little to understanding what actually limits performance.

Other Theories and Applications

In the following years, some attempts were made to counter criticism of resource models, by Gopher for one, mainly on the ground that their actual existence has been supported by findings in neuroscience. Attempts in the other direction were made to account for phenomena frequently observed by students of attention (such as the prevalence of performance decrements due to having to perform two tasks concurrently) by alternate theses.

One alternate thesis is the notion of outcome conflict, proposed by Navon and Jeff Miller, which claims that task interference is often due to a change in some alterant generated by a process that interferes with carrying out a concurrent task. Another notion that garnered wide popularity is a more moderate version of the single-channel hypothesis termed the single-bottleneck model, proposed by Harold Pashler, that assumes the existence of a central apparatus limited to executing only one process at a time.
and widely believed to be localized at the response selection stage. It was substantiated predominantly by findings in a paradigm of partly overlapping tasks called the *psychological refractory period* (PRP). Yet another notion, proposed by David Meyer and David Kieras, posits strategic postponement queuing, designed mainly to minimize outcome conflict. Later, Navon and Miller, and, independently, Michael Tombu and Pierre Jolicoeur, discussed the issue of empirically deciding between those notions and considered to what extent the controversy is decidable at all. Prominently, they argued that many PRP data may be also accommodated with divisible resource models.

Nonetheless, during the past two decades, it has become clear that, despite much criticism of resource theories, they still have their proponents, and even critics do not reject all conceivable hypotheses that posit limited internal input. Consequently, despite the manifest decline in enthusiasm for resource models, the notion of resources proved to die hard.

Because the notion of processing resources is a very prominent conjecture, many researchers have continued to use resource terminology, regardless of the extent to which it contributes, or not, to explaining and predicting behavioral phenomena, presumably out of linguistic preference (or habit) in much the same way that the computer metaphor is used. For example, many authors use the term to describe the deployment of visuospatial attention. Researchers in diverse fields of psychology—mainly cognitive aging and aphasiology but also child development—kept using resource models, sometimes in disregard of warned-of risks, such as vagueness, poor testability, dubious falsifiability, and circularity. For one, ascribing age-related cognitive deterioration to the depletion of the capacity of some resource pool may be no more than restating the observation by means of a term that appeals to a non-observable entity whose existence is questionable. It would require substantive corroborating evidence to count as a scientific explanation.

Yet, the methodological tools provided by resource theory continue to serve as a framework for empirical research of limits on performance.

David Navon

See also: Attention and Action; Automaticity; Multitasking and Human Performance; Psychological Refractory Period

---

**Further Readings**


---

**Attention and Action**

This entry deals with attention and its interrelations with goal-directed action. *Attention* is not easy to define. In the broadest sense, attention research concerns the neurocognitive processes by which coordinated, purposeful behavior emerges from the distributed activity of billions of neurons in the brain. As such, the field comprises some of the most challenging problems in science. This entry focuses on processes that enable sensory information, relevant to an organism's transient and/or long-term goals, to be prioritized for potential moment-by-moment control of action.

**Attention and Working Memory**

Most contemporary theories link attention with a similarly multifaceted concept of working memory. In these theories, working memory (WM) denotes the set of processes that, together, maintain a model
of the organism’s current environment, as related to its ongoing behavioral goals. Obviously, the content of WM needs to be continuously updated. However, this updating is highly selective. On this view, the content of WM is itself provided by (i.e., is the outcome of) attentional processes. At the same time, WM is seen also as the primary source of top-down attentional control, such that goal-relevant processing is prioritized (selected).

In most laboratory studies of attention, relevant and irrelevant stimuli (and actions) are specified by the experimenter. Outside the lab, however, behavioral relevance is determined by an organism’s entire goal hierarchy, with continuously varying priority accorded to unexpected threats and opportunities, as well as to the more orderly sequences of attention shifts needed for any skilled action. Ericsson and Kintsch argue that skilled human performance, from playing chess to making tea, requires not only information held in short-term WM but also rapid access to a vast set of retrieval structures in long-term memory. They call this long-term working memory.

Interest has centered on the role of various frontal (and parietal) brain structures, in particular the prefrontal cortex, in working memory functions. The emphasis here has been on the role of these brain structures in the active maintenance of patterns of activity representing current behavioral goals, and the means to achieve them. Influential models put forward by Desimone and Duncan, Miller and Cohen, and others propose that these activity patterns in the prefrontal cortex (and other structures, including the posterior parietal cortex) are the source of top-down bias signals, which serve to guide (i.e., selectively prioritize) the flow of neural activity throughout much of the brain. A brief summary of one of these models is provided in the next section.

**The Integrated Competition Hypothesis**

Desimone and Duncan’s integrated competition hypothesis rests on three general principles.

1. Neural activity within a cell population is (locally) mutually competitive. Thus, increased activation in cells responding to one stimulus is accompanied by reduced activation in neighboring cells responding to others.

2. Top-down priming of cell responsiveness biases this competition in favor of activations relevant to current behavioral goals. Thus, sensory neurons that encode a relevant attribute or location show both enhanced spontaneous firing rates and increased gain in stimulus-evoked responses. Top-down bias signals are carried by “backward” cortico-cortical connections. (Backward connections are those running from higher to lower levels, e.g., in perceptual systems, running in the direction of the sensory inputs.) Backward connections are in fact more numerous anatomically than the classical “forward” connections. Their role in contextual biasing of conflict-prone neural processing is fundamental.

3. Finally, local competition is integrated across widely distributed components of the processing system. Thus, as one focus of processing gains dominance within a given (e.g., sensory) cell population, this advantage is propagated to other parts of the network, where cells coding for other (e.g., motor) properties consistent with or associated with the same entity (including its location, reward-value, etc.) gain further support, so that the network as a whole tends to cascade into a state in which the neural representation of one consistent object-goal-action is dominant throughout. This property has been explored in many different connectionist network models. It suggests a basis for the broadly “one thing at a time” character of much of perception and perceptual-motor integration.

**Evaluation**

Within the cognitive-neuroscience community, the integrated competition hypothesis has rapidly become the most popular general framework for research on attention and working memory. It has gained support from a variety of sources, including single cell recording, functional brain imaging, network modeling, and behavioral experiments. As yet, however, the hypothesis has been applied only to a limited subset of attention-related behavioral tasks; in particular, some dual-task data are problematic. How complex human goals and goal-relevant actions are represented in frontal or parietal cortex remains essentially unknown; and figuring out how top-down biasing signals are directed, specifically, to enhance goal-relevant sensory and cognitive processing is another major challenge. Progress on both these fronts presumably requires an appropriate theory of unsupervised learning.
**Top-Down Priming, Predictive Coding, and Context**

One promising way of thinking about the top-down modulation of sensory processing, discussed by Friston, is in terms of generative or predictive coding. In generative models, the brain is basically viewed as an inferential (Bayesian) machine in which the backward projections from each unit are trying to predict their inputs at each level in the processing hierarchy. Learning processes in these models seek to minimize the predictive error by dynamically adjusting connection strengths, in such a way that the hierarchical structure of connections in the brain reflects more closely the hierarchical, context-dependent structure of the environment. (As an intuitive example, bird predicts beak, feathers, etc., feathers predict certain visual textures, shapes, etc.) In this way, higher levels in the brain provide contextual guidance to lower levels. Predictable inputs require little further processing or adjustment at higher levels. Unexpected inputs, on the contrary, generate a large error signal and thus are liable temporarily to “capture attention” in the network until their external cause can be appropriately encoded. A particularly interesting feature of these generative models is that top-down “attentional” modulation, bottom-up perceptual priming, and context effects in general, are all mediated by essentially the same functional architecture.

**Functional Integration and Segregation**

As outlined earlier, neural activity has an intrinsic tendency to propagate; that is, it is *integrative*. For example, activation of premotor units coding for a spatially directed movement (e.g., a hand or eye movement) co-activates sensory units coding for the same spatial target-location, thus giving those units a selective advantage; at a molar (functional-behavioral) level, that is to say, planning or preparation of a spatially directed action causes an “attentional spotlight” to be directed to the target location of the action. There is extensive evidence for such action-attention interactions, at both a physiological and a behavioral level.

An influential line of research initiated by Singer and colleagues concerns the synchronization of neural activity, observed across widely distributed cell populations. Synchronized firing is proposed as a temporary binding mechanism, linking together related activity in different parts of the brain. Evidence suggests that different synchronization frequencies might operate, similarly, to segregate processing between concurrently but independently active cell assemblies.

**New Perspectives**

These new ideas—and data—from the cognitive neurosciences offer a radically changed framework for understanding attention. Early students of attention, from the 1950s on, took for granted the idea that the flow of information processing in the brain was essentially unidirectional, from senses to effectors. Relationships between attention and action were similarly represented in terms of a one-way, linear sequence of processing stages: first the encoding of raw sensory features, then semantic encoding, followed by response selection and, finally, response execution. Selection (selective attention) was posited to intervene, as one further, obligatory step, either before or after semantic encoding. (The question was long debated.) As we now know, information processing in the brain is nothing like a one-way linear sequence of this kind. To the contrary, at every level, neural responses are found to be context sensitive; backward connections are at least as important as forward connections; goal- and action-related modulation of sensory activation occurs at all levels. In other words, there is no raw encoding, unmodulated by contextual, or semantic, relevance; there is no one stage or level for selection; and processing runs in multiple, parallel pathways, for example, encoding spatial versus nonspatial properties (“where” vs. “what”), and in many other specialized cognitive domains.

Early theories of attention, particularly the influential work of Broadbent, rested on the belief that selective attention was necessitated by a fundamental system constraint: the limited central processing capacity of the brain. The role of selective attention, Broadbent claimed, was uniquely to protect this limited-capacity system from informational overload. Broadbent’s hypothesis of a central limited-capacity channel had the merit of capturing the essentially integrative character of brain processing; however, his ideas about the role of selection can now be seen to transpose cause and effect. What is fundamental is the need for coordinated, goal-directed behavior. It is this coordination constraint that is the primary source of limited capacity in sensory-motor performance. The relevant causal process needed to satisfy this constraint concerns the moment-by-moment
prioritizing among competing, goal-directed actions, and thence—as selection-for-action—among competing sensory inputs.

**Processing Bottlenecks**

Theories of a global, limited-capacity channel or “workspace” continue to provoke controversy, as does the concept of one (or more) processing “bottleneck,” characterized by a strictly one-at-a-time (or “winner-take-all”) operation. Evidence for processing bottlenecks comes principally from two behavioral paradigms, the *attentional blink* (AB) and the *psychological refractory period* (PRP). Perhaps the most fundamental question about both sets of phenomena is the question of causation. Do these phenomena simply reflect inescapable system-limitations? Or are they the product—the behavioral signature—of integrative processes, of the kind briefly discussed earlier, but perhaps specific to the special conditions imposed?

The PRP bottleneck, interpreted as a fundamental system limitation on perceptual-motor performance, is challenged by studies of dual-task performance using quasi-continuous skilled tasks, such as reading aloud, sight-reading piano sheet music, “shadowing” auditory speech, copy-typing, and so on. Some task combinations (e.g., speech shadowing plus copy-typing) can show no detectable concurrency costs, even after minimal practice, whereas other combinations (e.g., reading aloud plus writing to dictation) result in near-total breakdown of one or the other task; furthermore, the same individual task can exhibit massive interference in one task combination and practically none in another. Proponents of a PRP bottleneck—as immutable system-limitation—assume that access to the bottleneck must be rapidly alternated between these concurrent tasks, making use of temporal chunking in the execution of speech, typing, musical performance, and other tasks. One obvious problem for this conjecture is that it fails to explain why alternation should work successfully in some task combinations but then fail completely in others. A deeper problem is that the conjecture appears to require a higher order control system, possessing intimate knowledge about the timing constraints in each task, in order to schedule conflict-free alternation. A more parsimonious interpretation might be that the special conditions under which PRP effects are obtained—discrete speeded responses to closely successive stimulus events, to be executed in a prespecified order—preclude the dynamic segregation (e.g., via neuronal synchronization) possible between sustained, structurally compatible tasks.

**Working Memory: Concluding Remarks**

The dual-task data suggest that functional segregation between ongoing, high-level perceptual-motor processes can be established, at least in some conditions. This finding is a challenge, likewise, for the integrated competition model: How, and/or in what circumstances, can the postulated, winner-take-all integration process be constrained to operate separately within each of two distinct processing streams? Dual-task performance also raises interesting questions for working memory. Tasks such as reading aloud, speech shadowing, and musical sight-reading make substantial demands on WM functions. (In reading aloud, for example, the eye typically runs many words ahead of the voice.) How does WM accommodate the maintenance of two such running representations, concurrently?

Research in the field of attention and action still has plenty of work to do.

*Alan Allport*

*See also* Attentional Blink Effect; Divided Attention and Memory; Psychological Refractory Period; Working Memory

**Further Readings**


ATTENTION AND CONSCIOUSNESS

The importance afforded to the topic of attention in cognitive psychology is partly owing to the fact that attention is thought to be intimately related to consciousness. This connection leads many psychologists to hope that theories of attention will enable us to make some progress toward understanding consciousness. How much progress is unclear. Some suggest that a theory of attention will tell us all there is to know about some of the mysteries surrounding consciousness. Others argue that, although attention and consciousness influence one another, they result from different neural processes, and so a theory of consciousness would remain to be given, even if a theory of attention were established.

Jamesian Roots of the Issues

In a passage that many psychologists cite as being the source of their basic conception of attention, William James (1890) suggests that attention and consciousness are closely related:

Everybody knows what attention is. It is the taking possession by the mind, in a clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. (p. 403)

In saying that the concentration of consciousness is the essence of attention, James indicates that there is a close link between attention and consciousness, but he does not seem to have thought that they are one and the same thing, as he took it to be possible for there to be conscious states that are not attention involving. Shortly after the passage just quoted, he describes one such “confused, dazed, scatterbrained state” in which conscious thinking apparently proceeds in the absence of attention. Even in this case, however, it is not clear that attention is wholly absent. James tells us that this “scatterbrained state” is the “real opposite” of attention, but he is also prepared to characterize it as a state in which attention is not absent but merely “dispersed so that the whole body is felt” (p. 404).

The point that James leaves unresolved continues to be pertinent in the current debates. Researchers who take the view that all consciousness involves attention are committed to explaining away cases in which we appear to be conscious but are not paying attention to anything in particular. In recent writings, Jesse Prinz has tentatively endorsed James’s suggestion that attention in such cases is present but dispersed. This response enables one to save the idea that everything consciously experienced is attended. But to make this response is to give up on the idea (which is equally Jamesian) that attention essentially involves selectivity and “focalization.” A proper assessment of this point will require further work on some neglected conceptual questions concerning the role of selectivity in attention.

Does Consciousness Require Attention?

The view that all consciousness is attention-involving is tempting for those who want to use theories of attention to inform our understanding of consciousness. It is, however, far from being an intuitively appealing view. The popularity it currently enjoys is due to several studies in which preventing attention from being given to a stimulus seems to result in that stimulus failing to be consciously registered. These studies take various forms, corresponding to the various ways in which a subject can be prevented from attending. In the simplest cases of “inattentitional blindness,” subjects are prevented from attending to a stimulus because they are performing a task that requires them to attend to something else for the entire duration of that stimulus’s presentation. In these conditions, subjects know very little about the stimuli to which no attention was paid. Such effects were first demonstrated for speech stimuli, under the rubric of the “cocktail party effect,” in the mid-20th century. They have recently been studied in detail in experiments using rapidly presented visual stimuli and in experiments employing more naturalistic stimuli, such as films of people playing simple games. In one such experiment, a surprising number of subjects who are giving their attention to players in a basketball-type game fail to notice the appearance of a pantomime gorilla.

Distraction by another task is not the only way in which attention can be kept away from a stimulus. Nor is it the only form of inattention that leads to an apparent lack of consciousness. In the “attentional blink” paradigm, a subject’s attention is kept from players in a basketball-type game fail to notice the appearance of a pantomime gorilla.

Disruption by another task is not the only way in which attention can be kept away from a stimulus. Nor is it the only form of inattention that leads to an apparent lack of consciousness. In the “attentional blink” paradigm, a subject’s attention is kept from a stimulus because the attention is taken up with the business of responding to a different stimulus, presented very shortly before. Stimuli that are presented within the period of this attentional blink seem not
to be consciously registered, as indicated by the fact that subjects fail to report them.

Other experimental effects in which subjects show a surprising lack of awareness of large stimuli are less obviously attention-involving, but they, too, have been interpreted as resulting from (and so as evidence for) a necessary link between inattention and a lack of conscious awareness. One such effect is the “change blindness” effect, in which subjects struggle to identify the one changing component in a pair of pictures that are repeatedly presented, one after the other, with a brief flash between them. Subjects in these change blindness experiments, unlike subjects in inattentional blindness studies or attentional blink studies, do not have their attention taken up with a distracting stimulus. Nonetheless, it is thought that, when these subjects fail to notice the differing elements in alternating pairs of pictures, it is because they have paid no attention to them.

Lack of conscious awareness has also been attributed to lack of attention in certain neurological conditions. Patients suffering with brain damage affecting the right parietal lobe often behave as if they were oblivious to items on the left side of space (a condition known as unilateral neglect). This condition is frequently taken to result from an inability to shift attention leftward. Interpreted in this way, the condition provides a further example in which the absence of attention seems to result in the absence of consciousness. It therefore provides some evidence for the claim that attention to an item is necessary for consciousness of it.

Ways to Defend the View That There Is Consciousness in the Absence of Attention

If we accept the claim that attention to a thing is necessary for consciousness of it, then we can account for the phenomena outlined previously, but to accept that claim is to accept that unattended background items never feature in conscious experience. That view seems unappealing from the point of view of common sense. Anyone who attempts to put this commonsense objection on firm empirical footing would, however, face methodological complications. To show that unattended items do figure in conscious experience, one would have to present subjects with stimuli that one was sure they were paying no attention to. It is unclear how this could be done. One would also have to show that the subjects were experiencing these stimuli consciously.

This would, problematically, involve ruling out the possibility that one’s evidence of consciousness could be explained away as unconsciously biased guessing. In the face of these complications, the argument against the idea that attention is necessary for consciousness has been driven more by theoretical considerations than by direct empirical demonstration. The emphasis has not been on showing that there are consciously experienced unattended stimuli, but on arguing against the reasons for thinking that there could not be such stimuli.

Defenders of the view that unattended items can figure in consciousness acknowledge that their opponents are able to make sense of the various inattentional blindness type effects detailed previously. But they can acknowledge this while maintaining that there are alternative interpretations of those effects that show the same explanatory advantages. There are two possible ways in which these alternative interpretations might be given. One might deny that the subjects in these experiments lack conscious experience of the unattended items (claiming instead that their failure to report these items results from lack of memory or from an incapacity to use their conscious experience as a basis for forming beliefs). One might, alternatively, admit that the subjects lack conscious experience but claim that something else, in addition to inattention, is responsible for this lack. Different cases may be treated in different ways. Unilateral neglect may plausibly be handled as a case in which the subjects’ lack of awareness is genuine but where factors other than inattention enable us to explain it. The case of inattentional blindness may be one in which the effect is due to lack of attention but where the unattended items do figure in conscious experience, although not in a way that has the structure, or the durability in memory, that would be required for the subjects to use this experience as a basis for forming beliefs: The effect of inattention might be understood as inattentional agnosia, or inattentional amnesia, rather than inattentional blindness.

Does Attention Require Consciousness?

In addition to the controversy about whether consciousness requires attention, there is a controversy about whether attention requires consciousness. Two lines of evidence suggest that it does not require consciousness, although, as before, there are routes available by which this evidence can be reinterpreted. The debate remains open.
The first line of evidence comes from observations of a blindsight patient. This patient lacks conscious experiences of the stimuli presented in one half of space. He nonetheless seems able to pay attention to those stimuli, as shown by the fact that cues indicating the location of these stimuli, whether presented in his scotoma or in his unaffected visual field, can, as attention directing cues would be expected to, increase the reliability and speed of the responses that he makes (as guesses) when asked about the stimuli of which he lack awareness.

The second line of evidence comes from normal subjects in conditions in which conscious experience of a stimulus is prevented because that stimulus is presented to just one eye, and the conscious experience of it is blocked by the presentation of more vivid stimuli to the other eye. When the stimuli that are presented in these conditions are erotic photographs, they attract and repel the subject’s attention (in ways that depend on the subject’s gender and sexual orientation) despite the fact that these erotic photographs are not consciously experienced. Subjects seem to be paying attention to these photographs but are not conscious of them. The example seems therefore to show that consciousness is not necessary for attention.

If facilitated processing is taken as an indicator of attention, then these studies indicate that a stimulus can exert an influence on attention without being a stimulus of which the subject is conscious. This is not quite enough to establish that a subject may pay attention to a stimulus of which he is not conscious. One might claim that the stimuli in these experiments influence attention (perhaps by drawing attention to their location) without the stimuli themselves being attended. The empirical results are, as before, amenable to alternative interpretations, with the consequence that the empirical argument is, at present, inconclusive.

Christopher Mole

See also Attentional Blink Effect; Blindsight; Change Blindness; Inattentional Blindness

Further Readings


**Attention and Emotion**

Our environment constantly confronts us with a large number of stimuli, some of which may pose threats and challenges that we need to respond to rapidly. Due to the capacity limits of the brain, only a subset of all the environmental stimuli that we encounter can be selected for more elaborate processing and subsequent access to other systems such as memory, motor control, and conscious awareness.

Several factors can have an impact on the competition for neural processing resources. Imagine you are searching for a specific object, for instance, your car keys. In this situation, perceiver-related factors such as expectations or voluntary behavioral goals can boost the neural representation of a stimulus via a top-down modulation of sensory pathways from frontoparietal regions. This voluntary selection control is referred to as *endogenous attention*. Conversely, stimulus-related factors, such as sudden and unexpected brightness or loudness, or stimulus uniqueness, may lead to enhanced responses in the sensory stream and trigger a reflexive orienting, which is referred to as *exogenous attention* capture.
Remarkably, the allocation of processing resources for perception and awareness also seems to be influenced by the emotional relevance of a stimulus. The adaptive function of emotions is to rapidly determine the value of a stimulus for well-being and survival and to coordinate appropriate behavioral responses. Emotionally significant stimuli should therefore be processed efficiently, even when they are not related to current voluntary goals, so that they can be noticed more readily and, once detected, become the focus of attention, evaluation, and action. This third selection mechanism, which may be designated as motivated or emotional attention, involves both exogenous and endogenous factors and is driven by specific neural mechanisms centered on the amygdala, a brain region critical for emotion processing and learning.

This entry will give an overview of how emotions can interact with attention to shape our perception and awareness of the environment, considering both behavioral studies and brain imaging evidence. The different kinds of effects that emotion can have on attention and perception will be illustrated, and the neural mechanisms that drive emotional attention will be discussed.

Behavioral Effects of Emotion on Perceptual Processing

Behavioral findings demonstrate a facilitation of perception and prioritization of attention allocation for emotional information. Emotional stimuli may capture attention more rapidly and/or impede attentional disengagement longer than neutral stimuli. Depending on the task and situation, this can improve performance when the target stimuli are emotional or lead to interference when an emotional stimulus competes with a non-emotional target.

In visual search tasks, the detection of a target among distractors is typically faster when the target is emotional as opposed to neutral. For example, people are faster to detect a spider among flowers than a hedgehog among flowers. Conversely, emotional distractors may impair search for a non-emotional target (people are faster to detect a flower among hedgehogs than a flower among spiders). However, more efficient search does not imply that emotional stimuli are processed without attention or visually “pop out” like targets defined by salient feature differences (e.g., color). Detection times are typically faster when targets are emotional rather than neutral but nevertheless increase when an emotional stimulus is surrounded by a larger number of distractors (people need more time to find a spider among ten flowers than among five flowers). Moreover, changes in task demands may sometimes affect the strength of such effects. Thus, attention appears to be preferentially guided toward emotional stimuli, reflecting biases in the allocation of attention rather than a direct shortcut to conscious perception.

In the attentional blink task, words are presented successively at a fixation (at about 10 Hz), and participants have to detect predefined targets. Detection is impaired when a target occurs shortly after another target. However, this attentional deficit is greatly attenuated for emotional target stimuli following a neutral target. Conversely, the deficit may increase for a neutral target following an emotional one, suggesting that the emotional meaning of items tend to capture attention in situations where resources cannot be equally deployed to every successive stimulus.

Emotional interference with performance is exemplified by variants of the Stroop task, in which participants name the color of either emotional or neutral words. Slower responses to emotional words are interpreted to reflect attentional interference by the emotional meaning. Usually, more interference is observed for negative than positive or neutral words, particularly in clinical populations with anxiety disorders.

When attention has been drawn toward the location of an emotional stimulus, it may also dwell longer at this location and facilitate the processing of subsequent stimuli. For example, in the dot probe task, participants respond to a target that replaces one of two simultaneous cues, one being emotional (e.g., a fearful face), the other one being neutral (e.g., a neutral face). Responses are usually faster toward targets replacing the emotional than the neutral cue. Emotional cueing may also increase contrast sensitivity for the subsequent target. These cueing effects are essentially exogenous and reflexive, as the emotional cues are not predictive of target location and hence do not help improve overall performance in the task.

Emotional orienting effects may arise with both negative and positive emotional cues. However,
some findings suggest that positive emotions or mood states may broaden the breadth of attention, whereas negative emotions primarily serve to focus attention. Emotional cueing can also operate across sensory modalities (i.e., for visual targets following auditory cues), suggesting that the prioritization of emotional stimuli is organized supramodally.

Although the behavioral results reviewed here strongly suggest that the emotional meaning or relevance of a stimulus may drive attention, one cannot exclude the possibility that some effects are due to more basic characteristics of the task or stimuli (possibly correlated with their meaning, such as the basic V-shaped form of the eyebrows in an angry face) rather than to genuine emotional appraisal. A role of emotional processing is however supported by the findings that attentional biases toward emotional stimuli can depend on the current emotional state or long-term emotional characteristics of the individual. For example, in visual search tasks, snake phobics are faster to detect pictures of snakes than of spiders, but spider phobics are faster to detect spiders than snakes. Depressive individuals rapidly shift their focus of attention toward negative information (e.g., words like gloomy or failure), whereas optimistic people show the opposite pattern and shift their attention toward positive information (words such as happy or success). Differences in current mood states may similarly influence attention toward positive or negative information. In the Stroop task, larger interference is obtained for emotional words related to individually relevant topics, as demonstrated for spider phobics, social phobics, rape victims, and post-traumatic stress disorder (PTSD) patients. This malleability strongly suggests that prioritized attention allocation is determined by an appraisal of the individual emotional relevance and not just a reflexive response to basic sensory features of a stimulus. However, the efficacy or the degree of “automaticity” of these emotional effects on attention is likely to vary depending on the nature and the complexity of the sensory cues conveying an emotional value and on the perceptual systems implicated. For example, personally familiar (hence emotionally relevant) information such as one’s own name or face may also capture attention in visual search or Stroop-like tasks, but only when these stimuli are rare and unaccompanied by multiple distractors; this finding suggests that their processing is more amenable to voluntary control or more sensitive to resource allocation. On the other hand, individual differences can also determine the effectiveness of conscious attentional control.

To sum up, behavioral data indicate that perceptual processing and stimulus awareness may be enhanced for emotional information under conditions of limited resources (when stimuli appear in cluttered scenes or rapid succession), although such enhancement may vary depending on both the task and the observer.

**Neural Mechanisms Underlying Emotional Attention**

Consistent with the behavioral effects that demonstrate increased perception of emotional information, brain imaging data from studies using functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) have consistently shown stronger neural responses of perceptual and attentional brain systems to a great variety of emotional stimuli as compared to neutral stimuli. Such effects have been observed for emotional scenes in the occipital visual cortex, emotional faces in the fusiform face area, emotional body movements in the extrastriate body area, and emotional prosody in the auditory cortex. The neural increases observed for an emotional stimulus (relative to neutral stimuli of the same category) are similar to those that would be produced when attention is voluntarily directed to a given stimulus presented among distractors (relative to when the same stimulus is unattended), but the emotional effects tend to occur irrespective of voluntary attention.

Imaging studies show that increased responses in visual or auditory cortex are also obtained for previously neutral stimuli after these stimuli have been repeatedly paired with electric shocks or loud noises. This provides strong evidence that the increases in neural activation to emotional stimuli are not (exclusively) due to intrinsic sensory features of the stimuli but are related to their emotional meaning. Similarly, research using electroencephalography (EEG) or magnetoencephalography (MEG) has revealed modulation by emotion at several stages of cortical processing, including both early, sensory-related processes (100–200 milliseconds post-onset) and later, more cognitive processes.
The amygdala is thought to play a critical role in these effects, not only for processing the emotional significance of stimuli but also for mediating the influence of emotion on perception. Several imaging studies have shown that increased cortical responses to emotional stimuli are correlated with amygdala responses: The more the amygdala responded to the emotional meaning of a visual stimulus, the greater was the activation of visual sensory areas to this stimulus. The amygdala has dense reciprocal connections with widespread regions in the cortex, including perceptual pathways as well as prefrontal regions, allowing not only rich sensory inputs to amygdala circuits but also direct feedback influences of the latter on sensory pathways. Such feedback can boost the processing of emotional stimuli and possibly enhance their subsequent storage in memory. Compelling evidence for amygdala feedback on cortical responses in humans was provided by the finding that patients with amygdala lesions show no differential activation of their intact visual cortex to emotional versus neutral faces.

In many cases, amygdala activation may arise prior to awareness or full attention for the emotion-eliciting stimuli and hence serve to redirect attentional resources to currently unattended but relevant information. In some cases, however, amygdala responses to subliminal stimuli might be reduced by brief presentation, high attentional competition, or certain personality characteristics such as low anxiety. Therefore, the exact conditions and pathways allowing for subliminal responses and their regulation remain to be clarified.

In addition to a direct boosting of neural representations via amygdala feedback loops, emotional stimuli can also produce indirect effects by modulating activity within the frontoparietal network mediating voluntary attention. In the dot probe task, greater fMRI activation is observed in this network, particularly in the intraparietal sulcus, when targets are preceded by a fearful face as compared with a neutral face; this finding is consistent with enhanced attentional orienting on such trials. In contrast, reduced activation in the intraparietal sulcus is observed when targets are presented in the ipsilateral visual field (i.e., the visual field on the same side) after an emotional face on the contralateral side (i.e., the visual field on the opposite side). This suggests that the intraparietal sulcus may become unresponsive to ipsilateral stimuli as a consequence of attention capture on the contralateral side. These neural effects corroborate behavioral findings that emotional stimuli may not only draw but also hold attention to their location. Extending these fMRI results, EEG recordings during the dot probe paradigm show higher amplitudes of the early (P1) visual evoked potential to targets replacing an emotional versus a neutral face, further demonstrating that emotional cues may bias spatial attention and enhance visual processing for subsequent stimuli that appear at the same location. Spatiotemporal analysis of EEG recordings also suggests that the visual enhancement of the target stimuli is preceded by a modulation of parietal activity, suggesting that parietal regions might be responsible for generating spatially selective top-down influences on visual cortex subsequent to the emotional cue.

The vast majority of imaging studies concerning emotional influences on perception have used negative or threat-related cues, consistent with physiological and evolutionary considerations suggesting a key function for the amygdala in fear processing. However, a similar enhancement of cortical responses has been reported for positive visual scenes, erotica, pictures of babies, or joyful voices, although their relation to voluntary attention mechanisms has been less systematically investigated than for negative stimuli. Because these positive stimuli also activate the amygdala, it is likely that emotional effects on attention are not exclusively sensitive to fear, but are more generally sensitive to emotionally relevant or arousing information. It remains to be determined whether a direct and reflexive activation of the amygdala and its feedback projections to sensory cortices might be more common and effective in response to simple threat cues (as fearful faces) than to other emotional stimulus types.

To sum up, interactions between attention and emotion can be observed not only behaviorally but also at the neural level, involving a direct boosting of stimulus representations in sensory cortices as well as a modulation of frontoparietal mechanisms responsible for orienting attention in space. These interactions rely on a large network that is centered on the amygdala but has many reciprocal connections with other regulatory systems in the brain (such as orbitofrontal and cingulate cortex, dorsolateral prefrontal cortex, and intraparietal areas),
which could modulate the gain or threshold of emotional responses. This network is well suited to appraise the emotional relevance of stimuli by integrating representations of intrinsic or learned affective value with complex situational factors related to goals, expectations, experience, or personality. Together, these emotional mechanisms complement endogenous and exogenous attentional systems that are known to select and organize sensory inputs based on voluntary goals and low-level physical salience, respectively, and can therefore be considered as a specialized neural system for emotional attention in the service of fast and adaptive response to highly self-relevant events. Taken together, the converging data from psychology and neuroscience highlight an idea previously held by some philosophers, that emotional processing does not just add affective “feelings” to the content of perception but has a direct impact on the content of perception and awareness.

Patrik Olivier Vuilleumier and Tobias Brosch

See also Attention, Neuroimaging Studies of; Attention and Emotions, Computational Perspectives

Further Readings


**Attention and Emotions, Computational Perspectives**

There is now a wealth of evidence that the deployment of attention is strongly modulated by emotional salience. This evidence has been provided by empirical studies using behavioral, imaging, and electroencephalographic approaches. However, only a handful of computational models of the interaction between attention and emotion exist. These models attempt to identify the mechanisms that underlie key empirical phenomena, by answering questions such as how the processing of a stimulus might be modulated by its emotional salience and, more specifically, how task-irrelevant emotional stimuli compete for attentional resources. Such models are the subject matter of this entry. The relevance of theories and models of attention and emotion in computer applications (in particular, in human-computer interaction) will also be considered.

Because of their link to neurophysiology, connectionist approaches will be highlighted. In addition, because there has been very little modeling work on other emotions, the entry will exclusively consider the effects of threatening stimuli. Models are classified according to the form of attention-emotion interaction that they posit: direct, indirect, or strategic.

**Direct Interference**

Andrew Mathews, Bundy Mackintosh, and Eamon Fulcher laid out the first variety of attention-emotion models. They proposed that all stimuli are rapidly evaluated to determine their level of threat, with this evaluation probably occurring in the amygdala. Then, due to the salience of threatening stimuli, the activation of such stimuli is directly enhanced. Consequently, relative to emotionally neutral stimuli, threatening stimuli are given a competitive advantage. Behaviorally, this manifests in two ways. First, when presented as distracting stimuli, threatening items will (surreptitiously) attract attention more than will neutral items and will thus impair performance more on an attended task. Second, when attended, threat stimuli will be responded to faster than will neutral stimuli. Initial findings seemed to confirm this pattern of behavior. For example, in the emotional Stroop task, participants were slower
to name the ink color in which a threatening word was written than they were the ink color of a neutral word, although strong effects were most evident in clinical populations, such as the clinically anxious.

Probably the most notable computational model of this kind was Gerald Matthews and Trevor Harley’s connectionist model of the emotional Stroop. Although the model was not directly wired up to exhibit this pattern, competition between stimuli emerged during learning, and a threat-monitoring unit gave threatening stimuli a positive bias in this competition.

**Indirect Interference**

John Taylor and Nickolaos Fragopanagos proposed a more complex realization of attention-emotion interactions in their neurophysiologically constrained neural network model. Although their model contains a pathway by which representations of threatening stimuli can be directly enhanced (via the amygdala), their model also proposes indirect effects of threatening stimuli. Specifically, a ventral emotion circuit, including the amygdala and orbital frontal cortex, is connected to more dorsal attention networks, including areas such as the dorsolateral prefrontal cortex and posterior parietal cortex. The reciprocal connections between these emotion and attention networks are inhibitory; thus, top-down attentional control can be disrupted by the detection of negative stimuli. As a result, performance can be worse on an attended negative stimulus than on an attended neutral stimulus, because activation of the ventral emotion network disrupts attentional focus on the negative stimulus itself. This aspect of the Taylor and Fragopanagos model is supported by recent behavioral data and contrasts it with the earlier direct models.

**Strategic Interference**

Brad Wyble, Dinkar Sharma, and Howard Bowman introduced a model of emotional interference that emphasizes longer term, more strategic effects of threatening stimuli. Thus, whereas all the previously discussed models assume that negative stimuli have their full effect within the trial in which they are presented, the Wyble, Sharma, and Bowman model focuses on cross-trial effects. This is consistent with recent reevaluations of emotional interference experiments, which suggest that, at least for linguistic stimuli, lingering effects are created.

The strategic-emotional interference of the Wyble, Sharma, and Bowman model has some similarities to the indirect aspects of the Taylor and Fragopanagos model, although, as previously discussed, the time courses are different. Specifically, in the Wyble, Sharma, and Bowman model, lingering negative emotional representations inhibit cognitive control processes. These processes focus attention on task-relevant stimuli. Thus, in response to the detection of a threatening stimulus, attentional focus is in fact broadened to encourage vigilance toward a diffuse profile of potential further threat stimuli. It is suggested also that reciprocal inhibition between the rostral-anterior cingulate cortex (which is believed to be responsive to emotional salience) and the dorsocaudal anterior cingulate cortex (which has a cognitive function) implements such an interaction between threat and cognitive control. This said, a number of the cognitive-emotional areas that have been proposed to be mutually inhibitory could potentially support such a mechanism.

**Applications**

Many computing applications, particularly in the area of human-computer interaction, would benefit from a better understanding of attention-emotion interactions. In particular, there is considerable interest in developing computer systems that are sensitive to the emotional and attentional state of the user. Such systems would sense the state of the human user, through, for example, facial expression recognition, measurement of physiological state, or direction of gaze, and then adjust presentation to be consistent with this cognitive state. A classic work on this topic is Rosalind Picard’s *Affective Computing*.

Research is also focusing on interacting with users via agents, robots, and avatars that express emotions. Clearly, to consider the efficacy of such systems, it is important to understand how emotional stimuli interact with attentional processing. All of the previously discussed computational models suggest that incidental emotional stimuli (especially if perceived to be threatening) have a cost for task-relevant processing. Consequently, emotion-expressing animated characters in computer interfaces may distract from central tasks.

*Howard Bowman*
Attentional Blink Effect

The attentional blink is the term used for the marked deficit in awareness of a relevant perceptual event when this event is presented shortly after another relevant perceptual event. It is typically measured using the rapid serial visual presentation paradigm illustrated in Figure 1A, although auditory presentations have also been used. Participants are asked to report two targets (here the two letters) embedded in a stream of distractors (here the digits), all presented at the same location at a rate of about 10 per second. The targets are referred to as T1 and T2, and the time between them (referred to as lag) is systematically varied. Figure 1B shows the typical result: T2 identification accuracy suffers when presented at short lags after T1, and recovers within about 500 milliseconds. This temporary dip in T2 performance, called the attentional blink, has become central to theorizing about the limitations of our cognitive processes. This entry will treat behavioral and neural evidence in relation to such theorizing.

Mental Processing Limitations

How much information the brain can handle at once has been a major question within the science of the mind. In 1980, John Duncan argued that the answer depends on the nature of the information. In the visual domain, information hitting our retina is first processed in parallel by the brain. The important distinction is between information that is relevant to an

---

**Figure 1** (A) Typical rapid serial visual presentation task in which the observer has to identify two letter targets in a stream of digit distractors, all running at 10 items per second. (B) Fictitious but typical pattern of results for T2, with good performance at lag 1 (at 100 milliseconds, “lag-1 sparing”), followed by marked dip in performance up to about 500 milliseconds (“attentional blink”).
observer (information that needs to be remembered or acted on, referred to as target information) and information that is irrelevant (referred to as distractor information). In Duncan’s view, targets are thought to be passed on from the initial parallel stage to a second, limited-capacity processing stage that enables awareness of the object. This stage works in a serial fashion, processing objects one by one. Distractors, however, are rejected within the first stage and do not make it to the second stage. As a consequence, the brain has problems when there are multiple relevant objects, since the second level can only deal comfortably with one object at a time. Some target information is therefore likely to be lost. It is not surprising that, when about a decade after Duncan scientists discovered the attentional blink, they soon explained the phenomenon in terms of limited-capacity resources. Specifically, these theories have stated that the attentional blink occurs because T1 occupies the second stage for up to 500 milliseconds. Indeed, there is evidence that the attentional blink varies with T1 processing difficulty, for instance, as manipulated by masking (the disruption of a stimulus by either an immediately preceding or a following stimulus), inter-item similarity, or cognitive load.

Limited Capacity or Selection?
The attentional blink is a profound and robust effect, but it is not universal. Several conditions have been found under which multiple targets can be processed within a short period of time. One important case is lag-1 sparing: When T2 is presented at the very first temporal position following T1, detection is often unimpaired. This is somewhat difficult to explain from standard two-stage theory, and several additional hypotheses have been developed (many assuming a combined episodic memory representation of the two targets). Performance on T2 also improves substantially when the distractors between T1 and T2 are replaced with a blank. Most remarkable is that the attentional blink virtually disappears when the intervening distractors are replaced by even more targets. The second stage should now be completely overwhelmed, yet performance improves. This has made scientists propose alternatives to two-stage limited-capacity theory. According to selection theories, the cause of the attentional blink lies not within the limited-capacity stage, but within the filtering or gating mechanism that performs the transition from the first to the second stage. The fact that performance on T2 is fine when there are no intervening distractors indicates a crucial role for those distractors. Selection theories, therefore, assume that distractors lead to active suppression (i.e., disruption or inhibition) of the perceptual input. The attentional blink is the direct consequence of this suppression. This means that as long as no distractor information is encountered, multiple targets can be processed at the same time. Thus, selection theory assumes a more important role for distractor rejection than originally conceived by Duncan and a larger second stage capacity than proposed by most attentional blink theories.

Brain Activity
Electro- and magneto-encephalography studies of the attentional blink have so far revealed little effect of T1 on signals that correspond to the initial visual processing of T2, including semantic processing. This is consistent with a first, unlimited-capacity parallel stage. From 270 milliseconds onward, however, some signals are weakened by the temporal vicinity of T1, most notably the positive deflection at 300 to 500 milliseconds referred to as the P3. It is thought that the P3 reflects working memory consolidation, response selection, or awareness—in other words, classic second stage processing. The P3 correlates positively with T2 performance, and, in some studies, negatively with T1 performance. The latter has been interpreted as evidence for limited-capacity theory (as it suggests a direct resource trade-off between T1 and T2), but selection theories can also accommodate this finding. Functional imaging studies suggest that the sources of these signals must be sought in a network involving frontal, parietal, and temporal brain areas. The exact functions of these areas remain difficult to assess because of the relatively low temporal resolution of imaging techniques and because of concomitant effects of masking and task difficulty. Using specific types of T1 and T2 (faces and scenes), one study has provided evidence that the temporal cortex is involved in classifying the input, even in the absence of conscious report. The frontal regions were only active for reported objects, suggesting a role in our limited awareness. The parietal cortex may then serve to engage attention on the targets and perhaps suppress the distractors.

Christian N. L. Olivers
See also Attention, Resource Models; Attention and Consciousness; Psychological Refractory Period; Visual Working Memory

Further Readings


Attitude Change

Attitudes consist of our evaluative reactions to the world around us. Whenever we contemplate the people, objects, places, and issues that we encounter (i.e., attitude objects), it is virtually inescapable that we evaluate them, viewing them as good or bad to some degree. When an attitude is positive, we are inclined to approach and engage with the attitude object, but when an attitude is negative, we are inclined to avoid or reject it. In this sense, our attitudes serve as basic guides for navigating daily life by providing a summary, bottom-line assessment of the often multifaceted stimuli with which we interact. In addition to organizing our basic understanding of the world, attitudes have been shown to serve other important psychological functions. Interpersonally shared attitudes provide an important basis for building and strengthening social ties (a social-adjustive function). Publicly voiced attitudes allow the individual to assert and reinforce personal or collective values (a value-expressive function). Attitudes can sometimes also allow individuals to cope effectively with anxiety and threatening realities (an ego-defensive function). Given the significant functions attitudes serve, psychologists have long asserted that understanding the forces that shape and modify attitudes is of central importance in predicting the cognitive, affective, and behavioral responses of individuals. The conditions under which attitudes can be changed, and the mechanisms through which such changes are effected, are increasingly well understood in light of advances in attitude theory and research.

Traditionally, attitude researchers studied these issues in terms of the explicit evaluative judgments individuals make about attitude objects. Such judgments have typically been captured via self-report questionnaires. However, abundant recent research has employed indirect means of assessing attitudinal reactions. For example, researchers have created various categorization tasks that allow for the assessment of evaluative carryover effects, in which the presentation of an attitude object is expected to influence the processing of a subsequently presented stimulus differentially, depending on whether a positive or negative evaluation of the attitude object has been automatically activated. These indirect methods of attitude measurement are often assumed to tap a more implicit kind of evaluation that is not necessarily isomorphic with the explicit evaluative judgments an individual might make. Although controversies continue to characterize scholarly discussions of the implicit attitude construct, there is a general consensus emerging that two types of attitudinal responses can be meaningfully distinguished: explicit, deliberated evaluative judgments versus immediate, automatic affective reactions that are neither deliberate nor logical in their fundamental nature. Automatic affective reactions occur merely by virtue of the activation of stored mental associations involving the attitude object. In contrast, evaluative judgments involve propositional reasoning centering on the validation (or rejection) of assertions about the attitude object. In addressing issues of attitude stability and change, it is important to consider both types of attitudinal response.

Malleability Versus Stability of Attitudes

Questions about the stability of attitudes have been of long-standing interest, and certain cases of long-term stability have been well established. One relevant phenomenon is “brand loyalty,” in which consumers form an enduring positive attitude toward a particular brand of products. Another domain where attitudinal stability has often been reported is political attitudes; for example, attitudes toward
political parties tend to show considerable stability across the life span. More generally, many mechanisms of resistance exist that minimize the likelihood that well-established attitudes will be changed by persuasive appeals. Nevertheless, changes in attitudes have been documented in countless empirical studies, so it is important to understand the conditions under which these changes are most likely to be evident and the psychological processes that are responsible for them.

**Characteristics of the Attitude**

Attitudes vary in how strongly held they are, and stability (or resistance to change) is often taken as a key indicator of attitude strength. Attitude strength is a multidimensional phenomenon, and the resistance to change seen in the case of strong attitudes may be explicable, at least in part, by other features that typically characterize strong attitudes. Strong attitudes often arise from extensive experience with and knowledge of the attitude object, in which case there is less leeway for new experiences to modify the well-established mental representations supporting the attitude. Attitudes that are resistant to change also tend to emerge when individuals are embedded in social networks characterized by homogeneity with respect to a given attitude. In addition, there is also evidence that many kinds of attitudes are genetically influenced, presumably by the effects of genes on relevant sensory and hormonal processes, intelligence, and temperament. Estimates of the degree of genetic heritability of a given attitude are positively associated with resistance to change. To the extent that genes exert a similar influence on relevant physiological processes over time, relative attitudinal stability would be expected. Conversely, attitudes that are not strongly influenced by genes, regularly reinforced by the social environment, or well supported by personal knowledge and experience are much more likely to be modified by new experiences.

Other attitude features are related to malleability. An attitude’s valence is one very basic factor that constrains its malleability. Positive attitudes imply approaching and engaging with an attitude object, and these new experiences always hold the potential to modify one’s evaluation of the object. In contrast, negative attitudes imply avoiding and disengaging, so new experiences are much less likely to occur. As a consequence, negative attitudes are typically less malleable than positive ones, at least when individuals’ own choices determine the extent of their interaction with the attitude object.

**Characteristics of the Attitude Holder**

Some dispositional characteristics are associated with variations in attitudinal stability. High self-monitoring, the tendency to monitor social expectations and regulate one’s behavior to meet those expectations, is associated with more malleable attitudes; low self-monitors, in contrast, tend to regulate their behavior by reference to personal values and standards, and they are more likely to have clear and stable attitudes as a result. Dogmatism is another personality trait that is commonly associated with stable attitudes; indeed, its very definition implies close-mindedness and resistance to change. There are also “person-by-situation” patterns to attitude malleability, as in the case of “repressers” versus “sensitizers.” Repressers are individuals who cope with anxiety by avoidance and defensiveness, and such people are likely to avoid exposing themselves to information that threatens their attitudes in sensitive, anxiety-related domains. In contrast, sensitizers cope by actively seeking more information, and thus they should be more subject to attitude change in these domains.

**Mechanisms of Attitude Change**

When attitude change does occur, several distinct mechanisms can operate. Four distinct processes are considered here.

**Contextualization**

The evaluative significance of a given attitude object can change dramatically across different contexts. For example, a bright orange life vest might be considered a fashion disaster on the runway, but it might be the most precious thing imaginable on a sinking boat. Research indicates that both automatic affective reactions and more deliberately considered evaluative judgments can readily change across contexts. These changes emerge to the extent that an attitude object is mentally represented in terms of multiple features that can be differentially salient in different contexts. When attitude objects are associated with a diverse knowledge base, not all of this knowledge will be activated on encountering the object. If different subsets get activated
in different contexts, the net evaluative implications may change.

**Learning**

Much research on attitude change has focused on learning processes. Two forms of learning have been of focal concern, one focusing on affect and the other on cognition.

**Associative Learning**

Research on evaluative conditioning proposes that attitudes toward a given object will tend to take on the same affective connotations as other stimuli with which the object is repeatedly paired. Evaluative conditioning effects have been documented in many attitudinal domains. Advertising represents one domain where evaluative conditioning is a strategy of choice; in many advertisements, consumers learn little about the product other than the fact that it is associated with beautiful people, exciting scenery, popular music, and so on. The evaluative conditioning strategy, by creating new mental associations, can have an influence on immediate, automatic affective reactions as well as evaluative judgments that are based on these reactions. Although there have been innumerable demonstrations of evaluative conditioning effects, the underlying nature of the phenomenon remains controversial.

**Propositional Appeals**

Early research on attitude change focused on learning new information about the attitude object, thereby changing the beliefs underlying one’s evaluation. Classic persuasion research addresses this process. It is now understood that learning new information from a persuasive appeal, per se, is not as important for attitude change as the self-generated propositional thoughts about the attitude object that are triggered by the appeal. Contemporary persuasion models focus on the amount and type of cognitive elaboration that is devoted to a persuasive appeal as key determinants of attitude change and on the subsequent stability of the newly acquired attitudinal position.

**Subliminal Influence**

The possibility of subliminal stimuli influencing one’s attitudes has been a topic of interest for decades, but, until recently, almost no evidence was available to document it. Although an abundance of research has confirmed that various momentary reactions can be influenced by subliminally presented stimuli, only a handful of studies has shown that attitudes can be meaningfully swayed by such stimuli. These studies have generally indicated that people are only likely to be influenced by subliminal stimuli when they are in a particularly receptive motivational state. For example, people who are thirsty may be receptive to subliminal messages about a beverage. In the domain of attitude formation, extensive evidence exists for the contention that repeated subliminal exposure to a novel stimulus can enhance the positivity of its evaluation by virtue of enhanced processing fluency when the object is supraliminally encountered.

**Self-Generated Attitude Change**

Attitudes can also change by processes of self-justification and self-perception. When individuals engage in behavior that is inconsistent with their attitudes, for whatever reason, the attitude may change to come in line with the behavior. Cognitive dissonance theory argues that this change occurs because of a drive for consistency; when no sufficient external justification for a discrepant behavior can be identified, an internal justification must be found, and often this involves revising one’s attitude (“I really wanted to do that”). Self-perception theory, in contrast, argues that no motivational drive is required to explain this type of attitude change; rather, individuals simply make inferences about their attitudes based on the implications of their behavior, much as an observer would do. Notably, both of these forms of attitude change rely on propositional reasoning. As a consequence, the changes are likely to be evident in deliberate evaluative judgments, but they are not necessarily seen in measures of the immediate, automatic affect generated by the attitude object.

Ultimately, whether attitudes are static or dynamic depends on a range of variables, but there are a variety of distinct psychological processes that can give rise to changes in both relatively automatic and more thoughtful attitudinal responses.

Galen V. Bodenhausen

See also Attitudes and Behavior; Cognitive Dissonance; Motivated Thinking; Persuasion; Political Psychology
Further Readings


\textbf{ATTITUDES AND BEHAVIOR}

The relationship between attitudes and behavior has long been a topic of central interest within social psychology. Early theoretical and methodological approaches to the study of attitudes generally assumed a strong association between attitudes and behavior. Such an assumption can be found in many early definitions of attitudes, which frequently defined the construct in terms of its presumed influence on behavior. For example, Gordon Allport’s classic definition of attitudes as “mental and neural states of readiness” that influence an individual’s responses to objects and situations related to that attitude clearly implies that attitudes should exert a directive influence on behavior. Likewise, early theories of attitude structure also assumed a strong attitude-behavior association. This point is perhaps best exemplified by the highly influential tripartite theory of attitude structure, which postulated that behavior was one of the three components comprising attitudes, along with affect and cognition. Even within the attitude measurement literature, early researchers often assumed strong attitude-behavior associations. For instance, many of the early indirect measures of attitudes relied on observable behaviors as indicators of the underlying attitude.

However, by the 1960s, some researchers had begun to raise serious doubts regarding the assumption of a strong relationship between attitudes and behavior. Perhaps most notably, in a highly influential review of 46 studies examining attitude-behavior associations, Allen Wicker concluded that attitudes and behavior correlated, on average, at only .15 and rarely correlated above .30. Thus, he challenged the notion that attitudes are useful predictors of behavior and indeed raised concerns regarding the fundamental utility of the attitude construct. His criticisms and those of others, which came to be known as the “attitude-behavior problem,” ultimately did not discourage the study of attitudes but rather served as an impetus to better understand why attitude-behavior correlations were often so low. The responses to this question took a number of forms, and each of these responses has, to some degree, continued to be an ongoing theme in attitude-behavior research.

\textbf{Methodological Issues in Attitude-Behavior Research}

One set of answers to the attitude-behavior problem was methodological in nature. First, some researchers disputed the accuracy of reviews of attitude-behavior studies finding weak associations, arguing that these reviews only examined a restricted set of studies and that a broader consideration of studies suggested stronger associations. Other explanations challenged the methodology of the studies on which the reviews were based. For instance, some researchers suggested that weak correlations might be a result of flaws in the attitude measures used in past studies (e.g., people may not have always honestly reported their attitudes). Perhaps the most influential methodological challenge, however, was the specificity-matching principle advanced by Martin Fishbein and Icek Ajzen. These researchers argued that attitudes were most likely to predict behavior when the attitude being examined was at the same level of specificity as the behavior in question. For instance, very general attitudes are good at predicting broad patterns of behavioral responding across
a wide range of behaviors but are relatively poor at predicting any single specific behavior. Likewise, a very specific attitude is good at predicting a very specific corresponding behavior but not general patterns of responding across many behaviors. They noted that many past studies of attitude-behavior consistency involved situations in which researchers measured attitudes at a very general level but attempted to predict a very specific behavior.

**Moderators of Attitude-Behavior Consistency**

One of the most varied and rich literatures to emerge in response to the attitude-behavior problem was the research literature exploring moderators of the strength of attitude-behavior associations. This approach is based on the premise that it may be misleading to characterize attitude-behavior associations as generally strong or weak, because conditions exist where such associations can be extremely strong or nearly nonexistent.

**Attitude Strength Properties**

Of the many categories of factors that have been identified to moderate the relationship between attitudes and behaviors, the most extensively explored has been various properties of attitudes. Collectively, these properties have frequently been referred to as dimensions or determinants of “attitude strength” and have been postulated to regulate when attitudes will have a strong impact on behavior versus when attitudes will have little directive influence on behavior. These properties include characteristics of the underlying cognitive structure of the attitude such as its accessibility (the ease with which an attitude is activated from memory), ambivalence (the extent to which the attitude is based on conflicting positive and negative reactions), and attitude-relevant knowledge (the amount of knowledge associated with the attitude). Other determinants of attitude strength found to moderate the strength of attitude-behavior associations are subjective beliefs one has about the attitude, such as the extent to which people are certain of their attitudes and the degree to which the attitude is personally important to them. Still other research has identified differences in the process by which the attitude is formed, such as whether the attitude is derived from direct or indirect experience and whether the attitude is arrived at as a result of careful thinking about attitude-relevant information or as a result of non-thoughtful cues as important moderators of attitude-behavior associations.

**Dispositional Characteristics and Contextual Factors**

Other research has attempted to explain variations in the strength of attitude-behavior associations not in terms of variations in qualities of the attitude but, instead, in terms of differences in dispositional characteristics of people or differences in contextual factors present at the time the attitude is measured or the behavior is performed. For example, some researchers have noted that some people, low self-monitors, tend to be internally focused when determining how they should behave (i.e., they look to internal guides to behavior such as attitudes and values). In contrast, other people are high self-monitors in that they look for external guides to their behavior, such as the actions of others. Thus, as one might guess, attitude-behavior associations are stronger for low-self monitors than high self-monitors. Likewise, situational factors can also alter when people are internally or externally focused when looking for guides to behavior. For instance, situational factors as simple as seeing oneself in a mirror can serve to make people more internally focused. Even when people focus on their attitudes, they also need to recognize that the attitude has relevance for the behavior in question. Situational factors can determine the likelihood that people recognize the relevance of their attitude for a given behavior.

**Current Theories**

A final general theme to emerge in attitude-behavior research has been to account for the intricacies of the attitude-behavior link by developing formal theories that specify the processes by which attitudes lead to behavior and that identify other constructs that play an important role in this process. Several influential theories of this sort have been advanced. The earliest such theory was the theory of reasoned action proposed by Fishbein and Ajzen. This theory begins with the premise that the single most proximal determinant of behavior is an intention (i.e., a conscious decision to perform a behavior). Intentions can thus be seen as the conscious motivation to engage in a behavior. Intentions, in turn, are assumed to be a result of two determinants: attitudes and subjective norms. Attitudes are people's
Attitudes and Behavior

81

general evaluations of the desirability of performing the behavior in question. Thus, the focal attitude in this theory is always specific to the behavior in question. For example, if a friend asks you to wash her car (the behavior), the attitude according to this theory is the attitude you have toward washing the car. This differs from a more global attitude that may be related to your feelings about the friend who asked the favor. Subjective norms refer to people’s perceptions regarding how they think important others would regard the desirability of their performing the behavior. In the present example, this would include your beliefs regarding how you think your friends and family would want you to react to the request to wash your friend’s car.

Because the reasoned action model can be difficult to apply to behaviors over which an individual does not have personal control, Ajzen proposed his theory of planned behavior. The central idea of this theory is that the intentions one has to perform a behavior depend not only on attitudes and subjective norms but also on the control one has over the behavior. As such, behavior can be viewed as a continuum of difficulty, with simple everyday tasks at one end and difficult, specialized, and skilled behaviors at the other end. If an individual feels she has control and ability to accomplish some behavior, she will show greater intent to engage in that behavior.

In an attempt to integrate findings that some behavior can be both deliberate and planned with intentions mediating attitudes and behavior, whereas other behaviors can be more spontaneous and thus influenced more automatically by attitudes, the motivation and opportunity as determinants (MODE) model was posited by Russell Fazio. With the MODE model, Fazio argues that both spontaneous and reasoned reactions can be explained by taking into account both motivation and opportunity. An individual is more likely to engage in planned or reasoned behaviors if he is motivated (say, by personal relevance) and has the opportunity (e.g., lots of time) to think about how he will respond (behave) to an event. In such situations, an individual is likely to behave in ways that align with a specific attitude. For instance, in the aforementioned example, if you had time to think about the car washing request, you would wash the car based on your feelings toward completing that action rather than your feelings toward the person asking the favor (or some other relevant global attitude). Thus, the MODE postulates that for highly deliberative behaviors, attitude-behavior consistency processes operate in a manner similar to what has been described in theories such as the theory of reasoned action.

However, when opportunity and/or motivation to carefully deliberate about a behavior is low, the MODE specifies a different attitude-behavior process. The MODE posulates that in these situations, the focal attitude will most likely be global in scope. For example, if you do not have much time to think about the request to wash your friend’s car, you may base your decision to wash the car on your global feelings about the friend who made the request. This is because forming attitudes takes considerable resources, and therefore, humans cannot possibly have attitudes about every specific behavior or task in which we have to engage. Thus, without the motivation or opportunity to make a planned decision, behavior is a result of spontaneous processes based on a global attitude rather than specific information. The MODE postulates that the global attitude exerts its influence on the behavior by biasing in an attitude-congruent manner perceptions of the target of the behavior and behavioral situation and that these biased perceptions in turn lead to the behavior.

Leandre R. Fabrigar and Maia S. Kredentser

See also Belief and Judgment; Persuasion

Further Readings


Attraction

Attraction is an important factor for the development of romantic relationships. Social psychology traditionally defines interpersonal attraction as a positive evaluation of a specific person that involves the behavioral tendency to approach the person, having positive feelings for the person, and holding positive beliefs about the person. Attraction has also been described as a desire to form a friendly or romantic relationship with a specific person. In both approaches, attraction is distinct from the related idea of attractiveness, the positive and desirable qualities that a person may possess. This entry will examine some of the many factors that influence attraction, such as physical qualities and interpersonal dynamics.

Most research on attraction has been done in North America and has focused on unacquainted peers. This research started in the 1960s with experiments on perceived similarity. In the mid-1980s, research began to be influenced by evolutionary perspectives and gender differences in mate preferences. These trends have continued, employing increasingly sophisticated methods (e.g., physiological and genetic measures) and making use of new developments (e.g., online matching and speed dating).

Perceiving a potential partner as physically attractive is a key component of attraction, particularly for romantic attractions. For example, a study of more than 2,000 participants in a speed-dating service found that the selection of potential partners for a future interaction was based primarily on physical appearance for both men and women. Other research has identified the physical characteristics that make a person attractive. Women judged men having large eyes, prominent cheekbones, a big smile, a large chin, and high-status clothing as highly attractive; men judged women having large eyes, prominent cheekbones, and a big smile as highly attractive. Similarly, men rated women who possess baby-like characteristics of big eyes, small nose, and small chin as highly attractive. Seemingly innocuous cues such as seeing the color red can also make women appear more attractive to men.

Physical attractiveness also relies on body shape. Men rate women with an “hourglass” figure as most attractive, particularly when women are not over or under weight. Women prefer a man with broad shoulders and narrow hips, but only in conjunction with high financial status. Generally, attraction is also more likely when partners have similar levels of physical attractiveness such that an attractive person is attracted to a more attractive partner. Here the mechanism is that everyone wants the most desirable partner they can get, so that in an open market of interactions, everyone ends up with others of about their own value.

A great deal of research on romantic attraction uses models from evolutionary psychology. The idea is that modern individuals have inherited tendencies to pursue strategies that lead to the successful production of offspring. Men have evolved to prefer signs of fertility such as youth and beauty; women, signs of protection and support of rearing of offspring such as status and financial resources; and everyone, personalities (kindness, honesty) that facilitate the maintenance of pair bonds through child rearing. This approach has generated some strikingly original research, such as studies related to women’s menstrual cycles. One found that females who were fertile and most likely to conceive preferred less feminine male faces, especially for short-term relationships; in contrast, women who were less likely to conceive preferred more feminine male faces. Another found that in the context of a short-term relationship, women prefer men with a more masculine, lower voice, especially when they are fertile. In yet another example, based on rating videotapes of men, women who were fertile especially preferred men who confronted others and were physically attractive.

Personality also matters. First, it affects perceptions of attractiveness: Discovering a person has positive personality characteristics makes the person appear more attractive; learning about a person’s negative characteristics makes the person appear less attractive. Second, attractive personalities are a
Attribution Theory

Fritz Heider, in his book *The Psychology of Interpersonal Relations*, theorized that people function like naive scientists to determine causes of behaviors; people’s thoughts about the causes of their own and others’ behaviors are termed *attributions*. This entry briefly traces the history of attribution research following Heider’s original work, including early cognitive theories that emphasized “rational” principles for making attributions; later cognitive theories that purported to describe how people actually make attributions, particularly with respect to ability and morality; and recent theorizing about the role of affect.

**Early Attribution Theories**

Following Heider’s lead, subsequent attribution researchers considered cognitive variables and processes. For example, suppose an observer wished to explain why Jill was impolite to Joe. Relevant considerations include whether Jill is impolite to others as well, whether others are impolite to Joe, and whether impoliteness arises when Jill and Joe get together. If Jill is impolite to others, this suggests that the behavior was due to Jill’s being an impolite person. If others are impolite to Joe, this suggests that there is something about Joe that provokes impolite behavior. If Jill is not generally impolite, and Joe is not typically the recipient of impolite behavior, but impoliteness arises when they are together, this
Attribution Theory suggests that the problem resides in the interaction between Jill and Joe. Other theorists have focused on social desirability. Because there is social pressure to perform socially desirable behaviors and avoid socially undesirable behaviors, socially desirable behaviors can be attributed to social pressure, whereas socially undesirable behaviors provide better clues to what the person is really like.

Attribution theorists also have considered “non-common” effects, which can be illustrated as follows. Suppose that Leslie chooses between two jobs; both jobs are near the ocean, near good restaurants, and pay equally well, but the first job is more interesting than the second job. Leslie’s choice of the first job over the second one could confidently be attributed to the work being interesting rather than to the ocean, restaurants, or pay.

In addition to the foregoing “rational” attribution processes, early experiments also suggested a less rational process where observers overestimate the degree to which behaviors tell us about the traits of the people who perform them. Much early research seemed to indicate that observers give insufficient weight to the situations in which behaviors are performed, thereby leading to stronger trait inferences than are justified; this effect has been termed correspondence bias or fundamental attribution error. Later research, however, demonstrated that these effects are due, at least in part, to methodological artifacts. The use of different instructions or different experimental paradigms has been shown to eliminate or even reverse the tendency to underestimate information about situations. At present, the pervasiveness of the phenomenon in the absence of methodological artifacts is unclear.

Later Cognitive Theories Pertaining to Ability and Morality

As attribution findings accumulated in the 1960s and 1970s, researchers began to suspect that perhaps the type of trait examined in particular experiments partially determines the results. To see why, consider ability and the premise that people without ability only can perform badly whereas people with ability can perform well or badly. It follows that a successful performance of a behavior indicates high ability (because a person without the ability could not have done it), whereas unsuccessful performance is ambiguous with respect to ability (because anyone can fail). In contrast, in the domain of morality, it is the negative behaviors that are diagnostic of traits because immoral behaviors only can be performed by immoral people, whereas moral behaviors can be performed by moral or immoral people.

These arguments were refined later by researchers who pointed out that not all ability or morality dimensions work in this way. Let us first consider ability and imagine that Jill successfully negotiates a difficult piano sonata, whereas Sarah makes a free-throw. It is unlikely that a bad piano player would have been able to succeed at the piano sonata, whereas it is reasonably likely that even a bad free-throw shooter would sometimes make a free-throw. Therefore, the successful piano performance is more indicative of good piano playing ability than the successful free-throw is of good free-throw shooting ability. The general principle here is that successes are considered to be more indicative of underlying ability if the likelihood that a person without the ability could succeed is low.

Morality works differently and is more a matter of definition than of the probability of performing behaviors. For example, suppose that Pete performs a dishonest behavior and Tom performs an unfriendly behavior. Well, then, honest people are defined as those who never (well, hardly ever) perform dishonest behaviors, whereas friendly people are defined less stringently (even friendly people are expected to perform unfriendly behaviors at times). Consequently, Pete’s dishonest behavior is likely to result in a strong trait attribution that Pete is a dishonest person, but Tom’s unfriendly behavior is likely to result in a much weaker trait attribution about Tom’s unfriendliness.

Affect and Attribution

One characteristic the foregoing attribution theories have in common is that they focus on cognitive processes with little emphasis on affect. Affect has received more attention in recent research, at least in the domain of morality, with recent theorizing that some types of immoral behaviors (e.g., dishonest behaviors) cause observers to feel more negative affect than do other types of immoral behaviors (e.g., unfriendly behaviors). In turn, negative affect endows behaviors with greater attributional weight. Thus, for example, a dishonest behavior induces
much negative affect, leading the observer to infer that the person who performed it must be dishonest, whereas an unfriendly behavior does not induce much negative affect and does not lead to a strong trait attribution of unfriendliness. Several experiments demonstrated not only that different types of immoral behaviors induce different amounts of negative affect but that, when experimental manipulations are performed to influence the level of negative affect induced by immoral behaviors, attributional weight is influenced similarly.

In summary, the attribution area has been exclusively cognitive for much of its history but with a recent increase of attention on affective processes. However, a comprehensive theory that completely specifies the interactions between cognitive and affective processes and their effects on attributions has yet to be formulated.

David Trafimow

See also Belief and Judgment; Causal Theories of Intentionality; Emotion and Moral Judgment; Rationality of Emotion; Social Cognition

Further Readings


Audition, Neural Basis

Audition, or the sense of hearing, is one of the most complex and least understood primary sensory systems in humans. Research on the neural mechanisms of how we perceive acoustic signals has converged from two extreme directions. The first is from the periphery toward the central structures, initiated largely by scientists from Bell Laboratories in their attempts to make a better telephone. The second direction is from the most complex perception, human speech, initiated largely by psychologists and linguists. In recent years, these two directions have blended together to provide a better understanding of acoustic processing in general. This entry will describe the basic building blocks from the cochlea, up through the brain stem and midbrain, and continue on to the cerebral cortex. The bulk of the studies are based on animal models, and therefore the topic of language processing will not be covered.

Peripheral Processing

Hearing in terrestrial vertebrates begins as sound waves strike the head and body. Sound waves can be thought to be composed of two parameters: frequency (pitch) and intensity (loudness). The outer ear, or pinna, plays a key role in shaping the acoustic stimulus as it enters the ear canal and travels to the tympanic membrane (eardrum). The unique shape of the pinna causes different frequencies of the sound to be reflected and resonated, causing some frequencies to be amplified and others to be dampened. This can result in a quite different sound that reaches the tympanic membrane compared to what was in the air. These cues are important for localizing sounds, particularly in elevation, and are the reason that sounds seem to originate from inside the head when using earphones, where the amplification and dampening does not occur.

The vibrations of the tympanic membrane are amplified by the three ossicles (small bones) of the inner ear, which then converts these vibrations into a traveling wave within the cochlea, much like a pebble dropped into a pond creates waves toward the shore. The cochlea is a spiral structure within the temporal bone that contains the mechanical and neural elements that allow transduction of the sound energy into neural signals. This is done by the traveling wave moving the basilar membrane, on which sit the sensory receptors, the hair cells. The basilar membrane is cleverly designed such that the motion is greatest for low frequency sounds at one end, and the location of the peak motion moves from one end to the other as the frequency increases. Thus, there are specific hair cells that will be activated the most for each particular frequency that the person or animal is sensitive to. There are two classes of hair cells: the outer hair cells and the inner hair cells. The inner hair cells provide the sensory information to the brain, whereas outer hair cells influence how the basilar membrane moves and do not provide any
sensory information. The inner hair cells release the neurotransmitter glutamate onto an afferent neuron called a spiral ganglion cell (SGC), so named because the cell body of this neuron is located in the spiral ganglion. The SGCs are the first to generate action potentials, and there are approximately 20 SGCs for each sensory hair cell. The axons of these afferents combine to form the auditory-vestibular, or eighth, nerve, which then enters the skull and makes contact with the cochlear nucleus of the brain stem.

Spiral ganglion cells have exquisite frequency selectivity in their response, due primarily to the functional characteristics of the basilar membrane and the hair cells providing input to the SGC. Thus, at very low stimulus intensities, the SGC will respond only to a very narrow range of frequencies. This range increases as the stimulus gets louder, making a V-shaped tuning function. As adjacent hair cells are located at slightly different locations on the basilar membrane, the frequency range of the corresponding SGC systematically shifts from the apex (low frequencies) to the base (high frequencies). This systematic shift is maintained throughout the ascending auditory system up to the level of the cerebral cortex (described later).

**Brain Stem Processing**

Each SGC synapses onto neurons in each of the three divisions of the cochlear nucleus located in the brain stem. These divisions are the dorsal cochlear nucleus, the anterior ventral cochlear nucleus, and the posterior ventral cochlear nucleus. At this point, the responses of the neurons are monaural, as they only have input from the hair cells of one cochlea. However, the cochlear nucleus neurons then project to the superior olivary complex (SOC), which consists of several nuclei whose neurons receive about the same number of synapses from the cochlear nucleus on both sides. Thus, as opposed to the visual and somatosensory systems where neurons in the subcortical areas respond only to the contralateral visual field or body surface, respectively, auditory information is combined from the two ears at the second central synapse located in the brain stem. The neurons of the SOC are also tuned to the two main cues of horizontal sound localization, the difference in time between the arrival of the sound in the two ears and the difference in the intensity of the sound in the two ears. The former arises as the sound has to travel a shorter distance (and therefore arrives sooner) to the near ear compared to the far ear unless the sound is from straight ahead or behind. The latter arises because of shadowing of the sound by the head and body making the sound quieter in the far ear.

**Midbrain Processing**

From the SOC, as well as the cochlear nucleus, neurons send axons to the nucleus of the lateral lemniscus and the inferior colliculus. The nucleus of the
lateral lemniscus is little studied, and it is currently unclear what functional role it plays in audition. The inferior colliculus, however, is an obligatory station along the ascending auditory pathway and as such is a “bottleneck” of processing. This structure, located in the midbrain, is also composed of several different subunits depending on the species, but all contain a “core” region. Neurons in this region, as with the preceding stations, are organized by their best frequency response. As this structure is three-dimensional, shaped somewhat like an egg, neurons along a single slice through this egg all have the same frequency selectivity, much like a stack of plates, with each plate responding to a specific narrow frequency range that is just above those in the plate above it and just below those of the plate underneath it. Neurons in the inferior colliculus retain the binaural processing selectivity that was seen in the SOC and further refine the acoustic signal before transmitting this information to the auditory thalamus.

Thalamic Processing
The auditory thalamus is the medial geniculate nucleus and can be divided into two broad regions: one that maintains the high frequency selectivity, known as the lemniscal thalamus, and one where the frequency selectivity is much broader, known as the non-lemniscal thalamus. These two regions have different cortical targets, with the lemniscal thalamus projecting primarily to the primary auditory cortex and other core regions and the non-lemniscal thalamus projecting primarily to the non-core regions. Auditory cortex itself, in primates, is composed of a core region (including the primary auditory cortex, or A1), which projects in turn to a “belt” region. The belt region is a series of cortical fields that are adjacent to the core and are heavily interconnected with their immediate neighbors. The lateral belt areas also project to the “parabelt” region in a similar manner, which is located just lateral to the belt areas. Thus, the core region is not directly connected with the parabelt but must transmit information to the belt cortical areas, which, in turn, project to the parabelt.

Cortical Processing
The cerebral cortex is implicated in all forms of auditory perception. In the absence of auditory cortex, it is not possible to perceive the location of the sound in contralateral space (i.e., in space on the opposite side from the lesion) or to identify complex acoustic signals. Deficits in sound localization have been noted in human patients; this can occur with a lesion on only one side. Such individuals are still able to identify what the sound is by using the intact hemisphere, so there is no impact on speech perception or their perception of “things” in the auditory world. In much rarer cases, lesions occur in the auditory cortex on both sides; these people become functionally deaf, a condition known as cortical deafness.

Spatial Perception
The most studied, and consequently best understood, function of the auditory cortex is in its role in sound localization. It is clear that there is not a topographic map of acoustic space in the cerebral cortex, as there is a map of the contralateral body surface in the somatosensory cortex or of the contralateral visual field in the visual cortex. Many neurons in the auditory cortex are selective for the spatial location of the stimulus, and in most but not all cases, the location that gives the greatest response is in contralateral space. Neurons in A1 have moderately spatially tuned neurons, but this spatial tuning is sharpened in the caudal (toward the back) belt fields and is broadened in the more rostral (toward the front) belt fields. This implies that acoustic space may be selectively processed in the more caudal regions of auditory cortex but not in a topographic manner. Populations of neurons, then, are most likely to encode the spatial location of a stimulus, with each neuron providing some information about where the stimulus likely is and, equally important, where the stimulus likely is not. Such population codes have been shown to have enough information in the caudal belt areas to account for sound localization performance across different intensity levels and may be a general form of cortical processing of many different stimulus attributes across different cortical areas.

Nonspatial Perception
The processing of nonspatial information has resulted in a cloudier answer. The first and foremost
problem is that it is not entirely clear what the auditory cortex would necessarily be attempting to encode. It is relatively easy to assume how space is represented, that is, in azimuth and elevation, but it is not as straightforward for nonspatial processing. In addition, it may be that spatial processing is also computationally easier, so that it can be accomplished earlier in the cortical pathway. Nonetheless, attempts have been made, largely using vocalizations as a complex stimulus. Such experiments indicate that neurons in the rostral belt regions are more selective for different vocalizations than neurons in the more caudal regions. This is not always seen, however, and in some studies there is very little difference in selectivity between different cortical areas. In addition, studies comparing the responses to forward vocalizations compared to the same vocalizations played backward do not indicate any specific selectivity for vocalization stimuli per se. Thus, although time-reversed vocalizations do not sound very much like vocalizations, they are equivalently able to activate cortical neurons as the much more common forward vocalizations. Attempts at using other classes of stimuli, such as amplitude-modulated or frequency-modulated sounds, also show a similar lack of specificity in different cortical areas, although some studies have suggested that the rostral area in the core region does have a differential response to more complex stimuli.

Summary
The neurophysiological basis of auditory percepts thus remains poorly understood. It is known that the cerebral cortex is necessary for these percepts and that auditory cortex is a complex of multiple cortical areas anatomically structured into serial and parallel pathways. Acoustic space is likely represented by the activity of populations of neurons in the more caudal parts of the auditory cortex, suggesting that nonspatial features of acoustic signals, such as speech, would be represented by the activity of populations of neurons in the more rostral parts of auditory cortex. This has yet to be clearly defined, although this possibility is currently under intense scrutiny.

Gregg H. Recanzone

See also Hearing; Speech Perception

Further Readings

AUDITORY MASKING
Auditory masking refers to the situation in which one sound reduces a listener’s ability to detect or discriminate another sound. We are all familiar with situations in which what we are trying to hear is obscured by the presence of other sounds. For example, we may have difficulty hearing a friend speak at a rock concert, or we may have difficulty hearing a knock on the door when using a vacuum cleaner. These are examples of auditory masking.

In the psychophysical literature, the sound that is masking another sound is called the masker, and the sound that a listener is trying to detect is called the signal. Masking is often quantified by determining the smallest detectable sound intensity (threshold intensity) for the signal in the presence of the masker or sometimes the masker intensity required to render the signal undetectable. Detection depends in complex ways on the relations between the physical characteristics of the masker and signal. Important parameters include sound intensity, presentation time, frequency, and binaural presentation (the way the sounds are presented to the two ears).

Parametric Effects

Sound Intensity
Weber’s law states that the smallest detectable change in a quantity is proportional to the value of that quantity before the change. Weber’s law is common across sensory systems and, for hearing, can be restated as the general rule that the smallest detectable change in sound intensity is approximately proportional to the baseline intensity. So, as masker intensity increases, so must the intensity of the signal at threshold. Sound intensity is usually expressed as sound level in decibels (dB), that is, on a logarithmic scale. Equal multiples in intensity correspond to
equal intervals in dB. Expressed in dB, the signal-to-masker ratio at threshold is roughly constant, having a value in ideal listening conditions of about –10 dB for maskers and signals with identical frequency composition. This means that we can just detect a signal with a level 10 dB less (10 times less intense) than the masker level when the two sounds occupy the same frequency region (although see section on frequency selectivity).

**Temporal Factors**

Masking is maximal when the masker and the signal are presented at the same time (simultaneous masking). However, masking may also be caused by a masker presented either just before (forward masking) or just after (backward masking) the signal. Backward masking is a weak effect in trained listeners and only causes an increase in the threshold within 20 milliseconds or so of the onset of the masker. Forward masking, however, can persist for over 100 milliseconds after the offset of the masker.

**Frequency Selectivity**

The basilar membrane in the cochlea separates out the different frequency components of sounds. Each place on the membrane is tuned to a particular “characteristic” frequency and, for a pure tone, sensitivity decreases as the frequency of the tone is moved away from the characteristic frequency. The frequency selective properties of the ear mean that masking is greatest when sounds occupy the same frequency region. When sounds occupy remote frequency regions, the sounds are separated on the basilar membrane, and little masking may occur. This enables us to “hear out” sounds of different frequencies played at the same time (e.g., different musical instruments in an orchestra).

Masking is one of the main techniques used by psychophysicists to measure frequency selectivity. An example is the psychophysical tuning curve. To measure a psychophysical tuning curve, the frequency and level of a pure-tone signal are fixed. A masking tone (or noise band) is varied in frequency. At each frequency, the level of the masker is found that is sufficient to mask the signal. It is assumed that listeners detect the signal by using excitation at the place on the basilar membrane with a characteristic frequency equal to the signal frequency. When the masker has a similar frequency to the signal, the masker level required is low, because the masker and signal are not separated on the basilar membrane. As the masker is moved away from the signal in frequency, the masker level needs to be increased. A sharp “V” shape indicates good frequency selectivity (i.e., good separation of frequencies) at the place tuned to the signal.

For complex sounds containing a range of frequency components, masking depends on the degree of overlap between the sounds in their cochlear representations. For example, a sound with a wide range of frequency components could mask a sound with a narrow range of components, but not vice versa. At high levels, low-frequency sounds can mask sounds of all frequencies. The rapid increase in the effectiveness of low-frequency maskers with increasing level is called the “upward spread” of masking.

Psychophysical data on frequency selectivity have been used as the basis for data compression in the MP3 sound format used in portable music players such as iPods. A sound that is masked does not need to be encoded accurately, and so the file size can be reduced.

**Binaural Factors**

The brain can separate the inputs to the two ears to some extent, so that little masking occurs
for maskers and signals presented to opposite ears, and masking is greatest when masker and signal are presented to the same ear. For the same masker and signal presented to both ears, threshold is reduced when the timing of the masker or signal is changed between the ears, for example, if the signal is inverted in one ear compared to the other while the masker is the same in both ears. This effect may be related to the neural mechanisms used to localize and separate sounds based on their time of arrival at the two ears.

Christopher J. Plack

See also Audition, Neural Basis; Hearing

Further Readings


AUTISM

Pervasive developmental disorders (PDDs) are a group of neurodevelopmental disorders with similar behavioral profiles and are referred to as autism spectrum disorders (ASDs). In this entry, when an autistic disorder is not specified, autism will mean ASDs. The entry will outline the current diagnostic criteria for autism, discuss the characteristics and course of the disorder, address the known pathophysiology, and provide a brief overview of assessment and treatment.

Current Diagnostic Criteria

The Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR), defines criteria for autistic disorder and PDD-not otherwise specified, as well as Rett’s disorder, childhood disintegrative disorder, and Asperger’s disorder. The diagnosis of autistic disorder in the DSM-IV-TR requires six symptoms, at least two of which are social, one communication, and one repetitive behavior/resistance to change. Rett’s disorder is a severe disorder of early development, involving deterioration of motor and cognitive skills, occurring primarily in girls, and associated with a specific gene. Asperger’s disorder shares the social and repetitive behaviors of autistic disorder but does not show early language delay or mental retardation. Childhood disintegrative disorder is a rare condition, in which severe autistic regression and poor outcome follows at least 2 years of normal development.

Diagnoses That Commonly Occur With Autism

Disorders that commonly occur with ASDs include mental retardation, anxiety, depression, tic disorders, and seizure disorders. Although the DSM-IV-TR does not allow for co-occurring diagnoses of autism and attention deficit hyperactivity disorder (ADHD), several recent studies demonstrate that ADHD symptoms are common in ASDs.

Characteristics

ASDs are characterized by an impaired ability to form normal relationships, deficits in the use of nonverbal forms of communication such as eye contact and gesture, lack of social reciprocity, poor awareness of others’ reactions, and failure to share experiences with others. The back-and-forth of conversations appears to be an area of especially profound pragmatic language deficit. Social cognition (processing of social information), especially in the area of processing face identity and facially expressed emotion, is also consistently impaired.

Children with an ASD generally present with significant impairment in language functioning. Although some autistic children are mute, verbal autistic children are typically, though not always, able to acquire normal morphology, syntax, and phonology. However, development of language is usually delayed, and pragmatic, abstract, and inferential language is weak.

An overly narrow focus of attention characterizes ASDs; this narrow focus may contribute to the phenomenon of savant skills (exceptional ability in one area) and lead to extreme degrees of perseveration, routine-boundedness, resistance to changes in the environment, difficulty with divided attention tasks, and difficulty in shifting attention. Sustained attention, however, appears to be spared, at least under certain conditions. A theory of “weak central coherence,” in which top-down processing is impaired relative to bottom-up and perceptually dominated
processing, has been proposed as an explanatory principle for the autism syndrome.

Executive functioning deficits (e.g., difficulties with organization, planning, flexibility, impulse control, working memory) are common in older children with an ASD. Younger children display such impairments less consistently. This suggests that such deficits do not explain core symptoms of ASDs, and are not universal in, or specific to, autism.

Developmental Course and Outcome
There appear to be two, fairly distinct onset patterns of ASDs: In the majority, development is delayed during the first and second years of life, especially in language and social development. In perhaps a third of the children, however, development appears to be essentially normal during the first 12 to 24 months, followed by regression of skills, especially language, and an abrupt development of autism symptoms. Lack of evidence for a specific environmental trigger for the regression has led to speculation of abnormal activation of innate immune processes as potential agents.

Autism diagnoses can be reliable as early as 18 to 24 months, and several groups are working toward identifying possible signs of an ASD in the first year of life. Studies of younger siblings of affected children, as well as home videos from the first year of life, suggests that failure to respond to name, emotional flatness, poor disengagement of attention, poor social orienting, and low rates of joint attention may be observable in children with an ASD by the end of the first year.

Behaviors may worsen in adolescence, and seizures sometimes begin. At the same time, adolescence represents a time of increasing interest in close relationships; although children with an ASD have difficulty developing these relationships, they may show increased social motivation at puberty.

Behavioral peculiarities and deficits in daily living skills may diminish as children mature and receive interventions. The most consistent predictors of good outcome in ASDs are IQ and the presence of early communicative language. In general, higher functioning children improve over time; a smaller percentage of individuals may even lose their diagnostic label. Recent research provides evidence for “recovery” in individuals with an ASD, and researchers are beginning to identify predictive factors of optimal outcome, as well as residual difficulties such as anxiety, depression, tics, and obsessiveness.

A substantial proportion of adults with an ASD need to live in residential facilities or with family members. The options for residential, semi-independent, and independent living situations are continuing to change and improve.

Epidemiology
Historically, autistic disorder has been thought to occur in 4 per 10,000 children; however, recent studies have reported higher prevalence rates. For preschoolers, recent estimates from the United Kingdom suggest rates of 16 per 10,000 for classic autism and 62 per 10,000 for ASDs. Debate persists about whether this reflects increased awareness and diagnostic changes, or a true increase.

Autistic disorder is more common in males than females by about 4 to 1. Females with autistic disorder tend to score lower on both verbal and nonverbal measures of intelligence.

Pathophysiology
The first major postmortem study of autistic brains by Tom Kemper and Margaret Bauman found small and densely packed cells in cerebellum, inferior olives, and limbic regions. Other studies, however, have found a heterogeneous group of brain abnormalities, with marked interindividual differences. Because social disability is prominent, studies of the amygdala have seemed promising but have yielded inconsistent evidence, and research suggests that it may be linked to anxiety rather than social dysfunction. Other functional magnetic resonance imaging studies point to abnormal processing of social information, especially faces, although abnormal brain activation during face processing may reflect impaired social attention, not impaired face processing per se.

Anatomical imaging studies have been inconsistent; Martha Herbert and colleagues found small and normal head circumference at birth, followed by acceleration of head circumference to a relatively large size by about 15 to 24 months, followed by slowed growth resulting in gradual return to average size, is one of the few replicated biological findings in autism. In a widely quoted study, Manuel
Casanova and colleagues reported that columns of cells in the neocortex (top layer of the brain) show abnormal structure. Theorists have suggested that white matter abnormalities result in overconnectivity between adjacent cortical areas and long-distance cortical underconnectivity, resulting in superiority in rote and concrete information processing but impaired abstract, integrative thinking; this is a dominant theoretical model of autism at present.

Genetic contributions to autism have been a focus of research since Susan Folstein and Michael Rutter identified a genetic contribution to its inheritance in 1978. Concordance rates are high, with pairs of siblings tending to have the same impairments or diagnoses. These rates are less than 100% in monozygotic twins, and recurrence rates in siblings are usually estimated at 2% to 10%. Environmental or epigenetic (changes in gene expression not due to changes in the DNA itself) factors could explain these rates. Only a minority of cases are linked to well-defined single gene or chromosomal disorders such as fragile X, whereas many, individually rare genes or chromosome regions have been identified. Several genes implicated in autism are related to components of the synapse. Some suggest that autism is a disorder of synaptic transmission.

Jonathan Sebat and colleagues reported that 10% of affected singletons had de novo (not seen in their parents) copy number variants in single stranded DNA loci (an unusual number of copies of a chromosome region without changes in the DNA making up that region). The implications of these findings are that copy number variation may be associated with autism, that singleton (one family member with an ASD) and multiplex (two or more first-degree relatives with an ASD) cases may differ genetically, and that singleton cases may reflect variation in a discouragingly broad array of possible loci. Clearly, the genetics of autism are extremely complex, but unraveling these complexities, particularly the transmission of singleton versus multiplex cases, will lead to more effective genetic counseling in the future.

Neurochemical research has focused on neurotransmitters, especially monoamines, glutamate, GABA (gamma-aminobutyric acid), and neuropeptides. The most replicated finding is that of peripheral hyperserotoninemia (elevated levels of serotonin in blood).

Assessment

Medical Assessment

In the absence of specific medical concerns, exhaustive medical workups for individuals with autism generally have limited clinical benefit; however, hearing should always be assessed. Blood screening for fragile X may be warranted. Epilepsy occurs in 11% to 39% of cases, and referral to neurology is recommended for any concern about possible seizures. Psychiatrists can assess comorbidity, including ADHD, depression, anxiety, and obsessive-compulsive disorder. If medication for medical conditions or severe behaviors (e.g., seizures, aggression, sleeplessness, depression, poor attention) needs to be considered, a psychiatrist or neurologist should manage the case.

Neuropsychological Assessment

Children with ASDs should have periodic neuropsychological assessments to describe the child’s current profile of cognitive abilities and behavioral challenges. Periodic reevaluations will help gauge the success of treatment programs and detect any deterioration of ability. An evaluation should assess language, visuospatial skills, abstract thinking and problem solving, memory, attention, and social cognition; in very young children, the assessment may be limited to the development of language, motor, and nonverbal reasoning skills.

Behavioral Assessment

A behavioral assessment is designed to guide treatment planning. A standardized instrument for the evaluation of adaptive skills in individuals with autism, such as the Vineland Adaptive Behavior Scales–II, will assess communication, daily living, socialization, and motor skills. Problem behaviors (e.g., social incapacity, resistance to change, self-injury, hyperactivity, aggressiveness) may require evaluation. In most cases, a functional behavior assessment by a behaviorally trained psychologist will be essential to optimal treatment planning.
**Family Assessment**

Parents and siblings of children with autism may experience psychological difficulties as a result of the stress associated with living with a child with autism, making at least an informal family assessment a critical part of treatment planning. Family assessments should address physical and emotional health of family members, as well as social support and family resources. Case management to obtain the best individual treatment, respite, and family support services from community agencies can be beneficial.

**Treatment**

*Pharmacologic Treatments*

As many as 70% of children with ASDs who are age 8 or older take psychoactive medication. Medication is better viewed as ameliorating specific associated symptoms rather than improving core language and social deficits. Antidepressants are sometimes helpful for anxiety, depression, sleep problems, and obsessive behaviors. Risperidone, an atypical antipsychotic, is approved by the U.S. Food and Drug Administration for children with autism ages 5 to 16 years who experience irritability, aggression, self-injury, tantrums, and mood swings. Stimulants have been used to treat inattention. Although some children benefit from taking stimulants, negative side effects including irritability, self-injury, social withdrawal, and insomnia may occur. Glutamatergic drugs and oxytocin hold promise as potential treatments for social and communication symptoms. Improvements in social behavior and relief from compulsive behaviors have been reported from use of oxytocin.

*Behavioral and Educational Treatments*

Evidence supports the use of applied behavior analysis in treating autism. Originally developed by O. Ivar Lovaas at the University of California, Los Angeles, in the 1960s, applied behavior analysis operates by breaking down skills to be taught (e.g., language, play, academics, social interaction) into very small steps, which are taught separately and then assembled into complex behaviors, using reinforcement principles of learning. Other behavioral approaches that incorporate naturalistic teaching approaches include the Denver model and pivotal response training. Structured teaching programs, such as TEACCH, are often used. Relationship-based intervention approaches are also utilized. A 2000 issue of the *Journal of Autism and Developmental Disorders* reviews treatments for autism.

“Alternative” treatments, such as “facilitated communication,” in which a facilitator assists the autistic individual to communicate by holding his or hand or arm while he or she points to letters, secretin infusions, special diets, megadose vitamins, heavy metal chelation, and others, have been shown to be ineffective.

*Deborah Fein, Colby Chlebowski, and Katherine E. Tyson*

*See also* Behavioral Therapy; Social Cognition

**Further Readings**


**Automatic Behavior**

This entry refers to a psychiatric and neurological state of mind, termed *automatism*, during which the person acts seemingly unaware of his actions, and examines the legal implications and consequences of these actions.
Automatisms are behaviors that happen outside conscious controls. Claims that an offender was not fully conscious at the time of the crime imply lack of cognitive scrutiny and will. Automatisms, along with acts committed by accident, under duress, compulsion, or threat, or from instinctively taking avoiding action, are not voluntary acts and are not punishable. The concept of free will as determinative of voluntary action is central to an understanding of automatism in law.

Automatisms result from medical conditions, external physical force, or serious emotional disturbances that may cause gaps of consciousness or interruptions in the thread of psychic life. During automatisms, a person may perform simple or complex actions, more or less uncoordinated, without being fully aware. Afterward, there may be confused memories or total amnesia for the episode. Acts that take place during an interruption in the psychic life of the individual qualify as automatisms. The use of automatism as a defense against a criminal indictment, although available in other latitudes, has been a concept mostly developed in the English-speaking world.

Habitual and mechanized acts that are repetitive and learned (gesturing, walking, riding a bike, driving, etc.) or “unconscious” acts in the psychoanalytic sense, such as Freudian slips, take place while the person is conscious; hence, memory for the event is preserved. These acts do not qualify as automatisms. However, deeply repressed materials may occasionally burst through in catathymic explosions, those uncontrollable emotional outbursts that are usually accompanied by dissociation. These acts qualify as automatisms.

Neurology and Psychiatry

A presumed association between epileptic seizures and violence was seminal to the development of case law on automatism. Offenders who commit a crime during an epileptic attack are usually found not criminally responsible. Automatisms are also observed in complex partial epileptic seizures characterized by motionless stare, interruptions of ongoing behaviors, stereotyped non-purposeful movements, and subsequent amnesia.

Dazed, semi-conscious victims of concussions may display automatisms. Other brain conditions also may lead to automatisms because the faster metabolic changes happen, either from normality or back toward normality, the more likely seizure is to occur.

Apart from automatisms observed in schizophrenia or during intoxications, three other conditions are associated with automatism in psychiatry:

1. **Dissociations** are coordinated physical acts undertaken while groups of mental processes or ideas are separated from the rest of the personality so that they assume an independent existence; thus, they are dissociated from the normal stream of consciousness. Dissociated acts escape the integrative activity of the mind and are common in hysterical neurosis, multiple personalities, autoscopy (the perception of self in the exterior environment) and in other out-of-body experiences, mystical phenomena, states of possession, voodoo, and experiences of mediums during séances in spiritualism.

2. **Psychological blows** are unexpected and sudden emotional shocks, on witnessing catastrophes or on receiving news, that seriously devastate emotional equanimity. A psychological blow defense presumes that any person, even one with intact personality structure and no history of mental problems, could be overcome and dissociate. Based on *R. v. Rabey*, courts apply an objective test for a psychological blow that measures the gravity and severity of the psychological trauma in a person with no history of mental or personality difficulties; absent this, courts will find for insane automatism, as in *R. v. K*.

3. **Somnambulism** is a state of dissociation in which, when the person sleeps, some fragment of the personality directs him into performing complicated acts, as asserted in *R. v. Tolson*. As determined in *R. v. Parks*, somnambulism is a “normal” form of automatism.

Law

Three factors lie at the intersection between medicine and law with regard to automatism as a defense:

1. **Memory.** Any one of the three memory functions, registration, retention, or recall, may be affected in automatism. Memory storage is not photographic; on recall, events are reconstructed according to life experiences, emotional needs, or unconscious needs. Recall may not be possible (amnesia) when events are not registered (as during intoxications), are lost (due to brain damage), or are
repressed (due to painful memories). Forgetting is a function of memory as is remembering. Amnesia, however, is not just forgetting but is a pathological process. Amnesia is crucial to an understanding of automatism; there is no automatism without amnesia, but not every case of amnesia is automatism. Determining the causes of amnesia—organic or functional, conscious attempts at distorting facts (lying)—and determining its timing, whether before or after an event, are important clues to understand a claim of automatism. Clinically and at litigation, it is essential to determine how far before and after the trauma the amnesia extends. Pre-traumatic amnesia (retrograde amnesia) affects an interval preceding the injury, for which the individual has no memory. Retrograde amnesia is short, usually less than 30 minutes; periods of weeks may signal hysterical manifestations. Anterograde amnesia develops after the trauma; it is longer and may last for months. Retrograde amnesia is not important at an automatism defense because the offense would have happened before the trauma, which itself could have caused the amnesia. Only anterograde amnesia counts for automatism.

2. Locus of cause. Causes of automatism are internal or external. A malfunctioning of the mind resulting from personal weaknesses or mental conditions is analogous to disease of the mind; these causes are deemed internal. Transient disturbances of consciousness such as concussions are not considered diseases of the mind, and, thus, they constitute external causes, as was determined in R. v. Quick. The internal/external test establishes the basis to decide whether an automatic act is insane automatism, meaning the act was caused by internal preexisting mental pathology, or sane automatism, meaning the act was due to external factors in the absence of preexisting mental problems.

3. Likelihood of repetition. The putative association between mental illness and unpredictable violent behavior underlies the need to determine whether an automatic act was sane or insane automatism. Internal factors due to mental conditions might repeat, but external factors such as a physical blow to the head or a psychological blow are unlikely to repeat.

A guilty mind, mens rea, is not a factor for a defense of automatism, but because an act, actus reus, is a crime only if it was willingly committed, absence of will is a factor for a defense of automatism. Automatisms negate the required mental state of voluntariness for commission of a crime; without voluntary control, there should be no guilty mind.

While the burden of proof for automatism rests on the accused, mere claims of its existence are not enough in courts of law. To have credibility, the claim requires “an air of reality” meaning that it must possess a certain degree of plausibility whose weight is to be determined at trial. If this test is met, then, the prosecution bears the burden of disproving the defense; reasonable doubts that the accused did not have the requisite volition lead to acquittal following an algorithm of unconscious act, no voluntary control, anterograde amnesia, external factor, unlikely repetition, and air of reality. Only unconscious, involuntary acts qualify for a defense of automatism.

Automatism is a total defense. Proof that a crime was committed during a state of insane automatism triggers application of dispositions for insane defendants and transfer to specialized forensic facilities. A finding of sane automatism leads to full acquittal. However, acquitting, on the basis of automatism, someone who has committed a serious offense raises multiple legal and sociopolitical concerns.

Julio Arboleda-Flórez

See also Amnesia; Automaticity; Autoscopic Phenomena; Consciousness and the Unconscious; Freedom of Action; Unconscious Emotions, Psychological Perspectives

Further Readings

Automaticity

Automaticity represents one of the oldest and most popular concepts in psychology. Although historically most popular in the context of cognitive psychology, the concept has made inroads into various
Automaticity has typically been conceptualized in terms of a set of criteria. Although there is some disagreement about the specific criteria, there is a modicum of consensus around the following three. An automatic process (a) does not require attention, (b) cannot be controlled, and (c) can occur outside of awareness. Each of these criteria is reviewed in turn, focusing on how each is assessed.

An Automatic Process Does Not Require Attention

Of all the criteria associated with automaticity, the idea that an automatic process does not require attention has probably attracted the most interest. This particular criterion has evolved to contain within it what might be considered sub-criteria associated with different forms of attention. For example, an automatic process should not require spatial attention, central attention, or any other form of attention (e.g., temporal attention).

The assessment of the attention criterion typically involves a manipulation designed to withhold a specific form of attention from contributing to the process of interest. If a given process is automatic, then it should be unaffected by a manipulation that prohibits attention from contributing to it. For example, research on spatial attention and automaticity has typically used manipulations that direct spatial attention toward or away from a stimulus. If the processing of that stimulus does not require spatial attention, then directing spatial attention away from it should not affect its processing. To illustrate, to test the claim that reading does not require spatial attention, spatial attention can be drawn away from a prime word (i.e., a word presented before a target word that is intended to prime target processing), and the impact of that prime word on the subsequent processing of a target word can be assessed. If word processing does not require spatial attention (i.e., is automatic), then the prime word should still influence processing of, and responses to, the target word when spatial attention is drawn away from it. Conversely, if word processing requires spatial attention (i.e., is not automatic), then drawing attention away from the prime word should eliminate its influence on target word processing. Research on other forms of attention (e.g., central attention) has used a similar logic but with different manipulations of attention.

An Automatic Process Cannot Be Controlled

The control criterion can be broken down into two distinct sub-criteria. First, automatic processes do not require intention. Intention can be defined in terms of an individual’s goals. An automatic process does not need an explicit goal or act of will to be initiated; it will occur simply in response to the proper stimulus.

The intentionality criterion may also be referred to as the volition criterion. An automatic process occurs involuntarily. Thus, an automatic process is said to be stimulus-driven as opposed to goal-driven. For example, if a word is present, then the processes involved in reading it are engaged without the requirement that the individual “intend” to read that word. The most widely accepted evidence that reading is automatic with regard to the intentionality criterion is the Stroop effect. In a typical Stroop experiment, the participant’s goal is to name the ink color of a word’s print (e.g., the word GREEN might be presented in red, and the correct response would be “red”). Despite this explicit goal or intent, the participant can often be shown to have read the word, even when reading the word is detrimental to performance (i.e., the printed word conflicts with
the ink color) and, therefore, is not likely intentional. A significant debate surrounds the conclusion that processing of the putatively unattended and irrelevant word in traditional Stroop tasks always takes place.

The second sub-criterion related to control is that automatic processes are **ballistic**. A process is ballistic if once started, it cannot be stopped. Thus, initiation of the process is a sufficient condition for the completion of that process. For example, suppose we assume that the activation of a word’s semantic representation is the “completion” of word recognition. Then, if word recognition is ballistic, a semantic representation will always be activated following initiation of the word recognition process. The stop-signal paradigm, in which participants are signaled to stop at different points in the execution of a task, has been used to test the ballisticity criterion. This criterion, which focuses on the control of a given process once it has been initiated, can be contrasted with the intentionality criterion, which focuses on the initiation of a process. It is possible to dissociate these sub-criteria, for example, by demonstrating either that process X does not require intention to begin, but that process X can be stopped midstream, or that process X requires an act of will to begin, but once started cannot be stopped.

**An Automatic Process Does Not Require Awareness**

An automatic process does not require conscious monitoring. Demonstrations that a process does not require awareness have typically relied on masked priming paradigms. In these paradigms, a stimulus is presented and masked so that the participant is unaware of what was presented. If stimulus processing does not require awareness, then eliminating awareness via masking should not inhibit processing of the masked stimulus. This paradigm has been used to determine whether semantic processing of a word requires awareness. When a target word (e.g., DOCTOR) is preceded by a related prime (e.g., NURSE), processing of the former is facilitated relative to when the target word is preceded by an unrelated prime (e.g., TABLE). By comparing conditions wherein participants are aware of the prime word and conditions in which participants are made unaware of the prime word via masking, researchers can determine if semantic processing of the prime word is dependent on awareness. If semantic processing is automatic in the sense that it does not require awareness of the word, then the semantic priming effect should remain even if participants are unaware of the word’s presentation. In contrast, if semantic processing is dependent on awareness of the word, then semantic priming should be eliminated when the mask renders the participants unaware of the prime. The conclusion from this literature appears to be that processing is not dependent on awareness. That said, there is still considerable debate as to what “awareness” is and how to measure it.

**Conceptual Issues**

The concept of automaticity has inspired more than its fair share of debate. One persistent area of debate surrounds which criteria should be used to identify a process as automatic. Three criteria have been described here. This list is not exhaustive, and different researchers assume different criteria. Other criteria that have been proposed include that (a) automatic processes capture attention, (b) automatic processes are fast, (c) automatic processes interfere with nonautomatic processes but are not interfered with by nonautomatic processes, (d) automatic processes do not require effort, and (e) automatic processes are independent of expectancy. The uncertainty surrounding the definition of an automatic process makes it critical that researchers be explicit about the criteria included in their definitions of automaticity. Unfortunately, many researchers fail to do this, and this failure leads to considerable confusion.

Another important issue in the conceptualization of automaticity concerns the extent to which a given process must meet all (stated) criteria in order to be considered automatic. Traditionally, an all-or-none definition has been used. Recently, however, a more piecemeal approach has been adopted, wherein a given process is considered automatic in terms of one criterion but not another. For example, “reading” may be automatic in the sense that it can occur outside awareness but not in the sense that it does not require attention. This lowering of the bar, so to speak, likely reflects the realization that no single process meets all the criteria for automaticity.

The last major conceptual issue surrounds the extent to which automaticity should be considered...
to be on a continuum or to be a discrete state. Although it is not controversial to claim that practice decreases the need for attention, control, or awareness (in whatever form), the idea that practice eventually leads to the elimination of the need for attention, control, or awareness is more controversial. This has led to the proposition that automaticity is better considered on a continuum rather than as a discrete state. In a continuous sense then, with practice, performance can become more automatic in that it requires less attention, becomes less controllable, and is less dependent on awareness, without requiring the claim that all need for attention, control, and awareness has been eliminated. Viewing automaticity on a continuum can be beneficial because it invites claims regarding the relative automaticity of different processes. For example, reading may be more automatic than color naming. A discrete view of automaticity, in contrast, does not allow for such comparisons. For example, if reading requires less attention than color naming, but both require some attention, then on a discrete view, both would be categorized as nonautomatic.

An Alternative Conceptualization

Dissatisfaction with the definition of automaticity led Gordon Logan to redefine the concept in his instance theory of automaticity. This theory differs in important ways from more traditional conceptions of automaticity. According to instance theory, performance on relatively unpracticed tasks is based on algorithmic processing, but with practice, performance becomes more reliant on direct memory retrieval. For example, the first time one is asked to multiply two numbers (e.g., $6 \times 4$), one might count upward by 6 four times to generate the correct response. The next time one is asked to multiply the same two numbers, however, the individual has some familiarity with that problem, and a response may be retrieved directly from memory. Thus, in this theory, automaticity represents the transition from algorithmic processing to direct memory retrieval. In this view, an automatic process is one that is completely reliant on direct memory retrieval. Both the storage of a new instance in memory and the retrieval of an instance from memory are obligatory but are dependent on attention. Thus, the instance theory diverges from traditional conceptions of automaticity.

Methodological Issues in the Assessment of Automaticity

In addition to controversy considering the conceptualization of automaticity, there has also been substantial controversy surrounding the assessment of whether a specific process is automatic. Much of this controversy has surrounded tests of the hypotheses that an automatic process does not require X (e.g., attention, awareness, intention, etc.). The assessment of this type of criterion requires a manipulation of that factor (e.g., attention, awareness, etc.) that is known to be effective. Unfortunately, ineffectiveness in the manipulation of the critical factor can lead to erroneous conclusions regarding the extent to which a given process requires that factor (e.g., attention, awareness, intention, etc.). For example, if a manipulation designed to remove attention from a stimulus is ineffective, then conclusions about the need for attention in the processing of that stimulus would be incorrect. Thus, if the presentation of a purportedly unattended word is shown to affect performance, one could conclude either that (a) reading does not require attention or (b) the manipulation used to control attention was not effective and the unattended word was inadvertently attended. Thus, the point of contention often concerns the determination of the effectiveness of these manipulations.

The persistence of these disagreements has left questions regarding the automaticity of specific processes unresolved, but recent work illustrating these controversies is resulting in more thoughtful analyses of experimental results, their implications, and theories regarding automaticity of cognitive processing. The field looks forward to an in-depth investigation of these issues.

Derek Besner, Jennifer A. Stolz, and Evan F. Risko

See also Automatic Behavior; Intentionality; Mental Effort; Stroop Effect; Unconscious Perception

Further Readings


Autoscopic Phenomena

Have you ever seen your own face in a flash-like vision, as if glancing into a suddenly exposed mirror—with the sole difference that there was no reflecting surface in the vicinity? Have you ever bumped into a figure, whom you first thought to be a stranger but who gradually made himself recognized, by an eerie feeling of psychological affinity, as nobody else than yourself, your “second self,” or doppelgänger? Have you ever felt separated from your own body, perhaps hovering a few meters above, looking down onto the body as if it were an empty shell you had just hatched from? As different as these scenarios may appear, they depict experiences subsumed under the heading of autoscopic phenomena.

Autoscopic phenomena form a heterogeneous class of reduplicative experiences of one’s own body and self. Literally, autoscopic means “seeing oneself,” but an autoscopic experience commonly involves far more than the visual senses. This entry describes the classification of autoscopic phenomena as currently accepted in medical and psychological literature. Based on phenomenology, this classification allows a rough interpretation of the different types of autoscopic experiences in terms of underlying cerebral mechanisms. The focus is then directed on the core type of autoscopic experiences, heautoscopy, which is the encounter of a person with a phantom double (i.e., doppelgänger) of herself. This multisensory, psychologically meaningful, and emotionally stirring experience has received much attention in anthropological and ethnological studies and in creative fiction as the “motif of the double.” The often profound symbolism contained in folkloristic and bellettristic accounts of heautoscopy is also found in clinical case reports and provides a challenge to a view of autoscopic phenomena that is solely based on neuropsychological mechanisms. Integrating formal aspects of an autoscopic episode and its symbolic content requires one to transcend current-day cognitive neurology. The study of autoscopic phenomena may thus contribute in important ways to bridging the gaps between neuroscientific views of body and self and the philosophy of the mind.

Phenomenology and Cognitive Neurology of Autoscopic Phenomena

As early as the end of the 19th century, French and German neuropsychiatrists recognized the need to distinguish several types of autoscopic phenomena. The classification they proposed is based on phenomenological criteria and has remained valid. In fact, the traditional classification scheme was only rediscovered in Anglo-American medical literature around the turn of the following century, with its typical urge to explain phenomenal experience in terms of underlying cerebral processes. Four major types of autoscopic phenomena are discussed here, along with the neuropsychological mechanisms currently assumed to be at the heart of the phenomena. Some rare types and related conditions are also mentioned.

Autoscopic Hallucinations:
Seeing an Image of Oneself

In an autoscopic hallucination, a person sees herself (often only the face or upper parts of the body) as reflected in a mirror. This specular form or “mirror hallucination” (in French, hallucination spéculaire) is almost always a consequence of brain damage, centered mainly around occipital or occipito-temporal areas. Whether the “extrastriate body area” is specifically involved is a matter of...
speculation. This cortical field is highly selective to the processing of images of body parts. The autoscopic image is frequently perceived in the blind parts of the visual field. Such hemianopic autoscopy is far more common in the left than the right visual field, which points to the dominant role of the right hemisphere for the recognition of faces in general and of one's own face in particular. In contrast to other forms of autoscopic reduplication there is not the faintest feeling of disembodiment; that is, the observer perspective is unequivocally centered on the patient's body. "Experimental" blinks with one eye are mirrored on the autoscopic face in the sense that right-eye blinks are seen as left-eye blinks and vice versa.

**Heautoscopy: Meeting One's Double (The Doppelgänger Experience)**

While autoscopic means "seeing oneself," the term heautoscopy denotes the "seeing of one's self" and thus implies the existence of two selves, both observers of one another. Hence, what is reduplicated in heautoscopy is more than a person's visual appearance. One's own doppelgänger "contains" one's self (or parts of it) and is a more lively and also more autonomous reflection than that in an autoscopic hallucination.

Visual features play a minor role in the experience. Reference to bright and natural colors is exceptional; one's phantom double is described, rather, as transparent, pale, or misty. In fact, it does not even need to look the same as oneself (in French, héautoscopie dissemblable). It is the psychological affinity with the doppelgänger that determines self-identification. This empathy with one's second self is accompanied, on the level of corporeal awareness, by shared bodily sensations. Increased ownership of the other is paralleled by feelings of depersonalization, of an abnormal lightness of one's body or sensations of "hollowness." Presumably, these complex illusions of bodily weight and spatial self-localization are mediated by the vestibular system and, anatomically, by the parietal lobes. In terms of embodiment and perspective taking, heautoscopy is halfway between the autoscopic hallucination (no feelings of disembodiment, body-centered perspective) and the out-of-body experience (disembodiment, perspective apparently centered on the reduplicated body).

**Figure 1  "Face to face"


Note: In heautoscopy, depicted in this illustration, a person encounters a phantom double, or doppelgänger. This is more than a mirror-like reduplication of one's visual appearance. In a heautoscopic episode, two selves are standing opposite each other, and the person is often at a loss to decide which of the two he really is. The long-standing motif of the double has inspired superstition, folklore, and creative fiction. In a clinical frame, one's doppelgänger's company may alleviate suffering but may also be experienced as an existential and unbearable threat.

**Out-of-Body Experiences: The Illusory Separation of Body and Self**

Core to an out-of-body experience is an abnormal localization of the self, more specifically, a disrupted
sense of spatial unity between body and self (disembodiment). The environment seems to be perceived from a location not coinciding with one’s physical eyes; visual impressions during an out-of-body experience are thus glimpses of the world from the mind’s eye. These glimpses may or may not contain an image of one’s body. If they do, this image is explicitly identified as one’s own body and not as a second self or doppelgänger. Out-of-body experiences are reported by persons in life-threatening situations and are frequently experienced by individuals with neurological disorders such as epilepsy and migraine. The neurophysiological mechanism is thought to be a transient disintegration of vision, touch, and proprioception in multisensory cortical areas at the junction between occipital, temporal, and parietal lobes. Especially in cases of epilepsy, this disintegration can wax and wane, such that a person may experience a kind of “double consciousness” or an intermediate state between heautoscopy and an out-of-body experience.

**The Feeling of a Presence: The Invisible Double**

This is the distinct awareness of another being in the near extracorporeal space. It is comparable to the phantom limb phenomenon in that the location and extension of the felt presence can be precisely pinpointed in space, and despite the absence of any visual confirmation, the feeling is overwhelmingly real. The “other being” is an invisible doppelgänger, even if the presence is rarely consciously identified as a second self. Self-identification is often inferred, however, from the observation that the nearby presence is following one’s own steps, mimicking one’s actions, or knowing one’s thoughts. Frequently observed in schizophrenia, the feeling of a presence is also reported by neurological patients with lesions of a predominantly parietal localization and tends to co-occur with out-of-body experiences and heautoscopy. Correspondingly, these three types of autoscopic phenomena can be induced by electrical stimulation of similar sites within the parietal cortex. The feeling of a presence is also known by mountaineers and explorers, particularly from episodes of exhaustion, lack of oxygen, and social deprivation.

**Rare Types of Autoscopic Phenomena and Related Conditions**

Negative heautoscopy (also asomatopy) refers to the not-seeing of one’s own body, either on reflecting surfaces or when looked at directly. Usually, this negative visual hallucination is accompanied by a similar failure in the somesthetic domain, that is, the non-awareness of having a body (asomatognosia). Inner heautoscopy (in French, *héautoscopie interne*) involves the visualization of inner organs in extracorporeal space. Many authors are reluctant to consider this experience an autoscopic phenomenon; it was frequently described in the early 20th century and almost exclusively in patients with hysterical disorders. All types of autoscopic phenomena discussed in this entry have been reported as poliopic variants, that is, a multiplication rather than a duplication of one’s face, body, and/or self. Poliopic heautoscopy can be conceived of as a whole-body variant of the illusory multiplication of body parts or supernumerary phantom limbs. Whether children’s imaginary companions, vivid autobiographical memories with visualization of oneself as an agent, or certain hypnagogic hallucinations should be brought into the conceptual vicinity to autoscopic phenomena is debatable. The syndromes of delusional misidentification do not have an autoscopic nature; the term *double* in those contexts refers to an impostor of another person or object, whose existence is inferred rather than perceived.

**Beyond Neurology: The Reflective Mirror of Heautoscopy**

Phantasmal reflections of one’s own body and self have long stimulated folk-psychological and academic imagination about the unity of mind, the multiplicity of personality, and the origin of the soul. Out-of-body experiences, cross-culturally largely invariant, are at the heart of dualistic notions about body and soul, and continue to nourish beliefs in an afterlife. On the other hand, in many cultures doppelgängers have the status of a harbinger of death—a justified view, perhaps, given the statistically short survival times after stroke, brain tumors, and other serious neurological diseases. The popularity of the motif of the double in creative fiction, peaking in Romantic literature, is undoubtedly due to the phenomenological varieties of heautoscopy. The essence of this particular type of autoscopic experience is best captured in the words of French artist Jean Cocteau (1889–1963), who once demanded that “mirrors should reflect a little more before throwing back images.” In fact, heautoscopic mirrors defy the laws
of optics. They are cognitively reflective, in Cocteau’s sense, in that they confront a person with a highly integrated, embodied image of the self. Handedness is maintained, which gives the image a fair portion of autonomy and the potential to develop a will of its own. In creative writing, this will is often contrary to a protagonist’s intentions, and the emerging antagonism classically culminates in killing the doppelgänger, that is, symbolically killing oneself—lively examples can be found in the work of Oscar Wilde (The Picture of Dorian Gray), Edgar Allan Poe (William Wilson), and Guy de Maupassant (Le Horla). A similarly dramatic hostility can be found in clinical case reports, albeit there are also benevolent doppelgängers, especially in cases of anosognosia (the nonrecognition or active denial of one’s own illness) where a patient may pity his second self for the bad shape he fails to recognize in himself.

Autoscopic phenomena thus form a set of human experiences formidably suited to illuminate the borderlands of neurology, psychiatry, and philosophy. Their study spans from visual self-recognition to the localization of the self in space and finally to issues of perspective taking in a double sense, visuospatially and psychologically.

Peter Brugger

See also Anosognosia; Capgras Delusion; Consciousness and Embodiment; Self-Consciousness

Further Readings


**Availability Heuristic**

This entry provides a brief explanation and background for the availability heuristic, a cognitive process hypothesized to underlie people’s intuitive judgments of probability.

**Heuristics and Biases**

The availability heuristic is a theoretical construct that forms part of the influential heuristics and biases framework for explaining intuitive judgment in humans, pioneered by Daniel Kahneman and Amos Tversky in the 1970s. From this perspective, because of their limited time, knowledge, and computational ability, in general, people cannot make judgments according to the often complex normative rules of logic, probability theory, and statistics. Instead, they have to resort to simpler judgment heuristics that exploit natural assessments. Natural assessments benefit from naturally existing propensities of memory and perception that are conveniently available and easily assessed. These judgmental heuristics are often useful, but sometimes they lead to systematic and serious cognitive biases.

With regard to confidence (probability) judgment, the key notion has been variable substitution, according to which the complex judgment of probability is substituted with a simpler natural assessment. Assessment of probability in the sense implied by probability theory and statistics involves consideration of all possible outcomes and their frequencies of occurrence. In the face of this complexity, it is proposed that people substitute probability with a subjective variable that is conveniently available and easier to assess. The availability heuristic, specifically, suggests that the ease with which an
event can be brought to mind or imagined is used as a proxy for probability, as when you assess the probability of an event by considering how easy it is to retrieve similar events from memory. One could estimate, for example, the risk of dying from pneumonia by assessing the ease with which one can come up with examples of people that have died from pneumonia. If it is difficult to come up with such examples, the risk of dying from pneumonia is presumably fairly low.

Although availability may be a useful guide to probability in many real-life circumstances, because the ease of retrieval or imagination is often affected also by factors other than probability or frequency, use of the heuristic may produce a number of biases, or cognitive illusions, in probability judgment. These biases may arise in two slightly different ways, either as a consequence of biases in the external flow of information itself or because of the properties of encoding and retrieval relevant to human memory.

**External Origins of Availability Bias**

The reliance on availability may in part produce biased probability or frequency judgments because of biases in the external information flow a person encounters in his or her environment, which make the ease of retrieval a misleading cue to the population frequencies. For example, people’s perception of the frequency of different death causes correlates more strongly with the frequencies reported in media than with the true fatality frequencies. People thus tend to overestimate the rate of violent and dramatic deaths that receive the most attention by the media (e.g., murder, car accidents) whereas they underestimate the risk of more mundane causes of death (e.g., pneumonia, cardiovascular disease). One may overestimate the prevalence of one’s own opinions and values in the population, simply because one tends to socialize specifically with people who have similar opinions and values. In a variety of domains, people are exposed to selective information and to the extent that people are unaware of this fact, reliance on availability yields biased judgments.

**Internal Origins of Availability Bias**

A second, perhaps psychologically more interesting, origin of bias derives from the operating principles of the human mind. These biases may derive either from the processes at the encoding of information into memory or from processes at the retrieval of the information. Events that are strongly emotionally charged or inherently interesting in other respects tend to receive more attention and therefore to become better encoded in memory. If you, for example, read a list of names of people, half of which are famous women (e.g., Madonna), the other half unknown males, afterward you are likely to overestimate the percentage of female names on the list, simply because they receive more attention and become better encoded in memory. Other judgment biases arise from the way memory is searched at retrieval. For example, in experiments in which people are asked if there are more English words starting with the letter r or more words that have the letter r in third place, people erroneously tend to guess that there are more words that start with r. The reason seems to be that we find it easier to search memory for words starting with a letter than for words with letters in third place; therefore, the first category is more “available” and appears to be more frequent.

It is clear that people’s judgments are often affected by the general sort of biases implied by the availability heuristic. Recent research has concentrated on determining the relative importance of putative external versus cognitive origins of the “availability biases,” if the judgments are driven primarily by the content of the retrieved material or specifically by the “ease” or fluency of the retrieval experience, and efforts have been made to relate the availability heuristic to more detailed computational models of memory.

Peter Juslin

See also Belief and Judgment; Debiasing; Representativeness Heuristic; Similarity

**Further Readings**


Behavioral Therapy

As the development of behavioral therapy has always been closely related to new evolutions in psychological research, it is hard to give a concise and definite definition. Nevertheless, throughout the past century, this therapeutic approach has become the most practiced, most researched, and best evidenced strategy to treat psychopathology in both adults and children. This entry will describe how current behavioral therapy developed in three waves. For each wave, the entry will describe the theoretical foundation, some examples of related therapeutic techniques, and some important points of discussion.

Next to other important research findings, the discovery of two learning principles has been extremely important for the development of behavioral therapy. On the one hand, Ivan Pavlov demonstrated that dogs learn that certain stimuli in a context (footsteps or a sound signal) function as predictors for food serving, which lead them, after some trials, to produce saliva in anticipation of food. This research has led to an increase in understanding of so-called classical conditioning principles. These learning experiences lead an organism to ascribe meaning to a context-stimulus. On the other hand, B. F. Skinner demonstrated that pigeons learn to press a lever if this behavior is followed by receiving a food pellet. This led to an increased understanding of so-called operant learning principles. These learning experiences lead an organism to increasingly show a certain behavior if that behavior is followed by a pleasant consequence (or reinforcement). If behavior is followed by an unpleasant consequence or by no reinforcement, its frequency will decrease until that specific behavior has been extinguished, which means that the behavior has disappeared.

Wave 1: Behavioral Therapy

Based on John B. Watson’s behaviorist principle that psychology should only research overt behavior, behavioral therapy originated in three different and independent locations, guided by Joseph Wolpe in South Africa, Hans Eysenck in the United Kingdom, and Skinner in the United States. Mainly following the principles of classical conditioning and operant learning, these pioneers approached maladaptive behavior as a reaction to a specific context (so-called stimulus-response [S-R] chains) in which this reaction should be replaced by more adaptive behavior. They deliberately objected to taking into account the non-observable inner world of the client and thus opposed difficult-to-investigate classic psychoanalytic claims that maladaptive behavior stems from unconscious processes.

Behavior therapists abstract a thorough so-called functional analysis from the clients’ description of their maladaptive behavior and the contexts in which they demonstrate this behavior. This functional analysis contains the relevant context-related stimuli, the behaviors, and the consequences of these behaviors and leads therapists to understand the forces that influence the provoking/eliciting or continuation of maladaptive behavior. Several techniques have been proposed that can be adopted to change the maladaptive behavior. Therapists can teach clients to
avoid exposing themselves to the stimuli that elicit maladaptive behavior (e.g., teach smokers to avoid places where many people smoke). Therapists can suggest that clients adopt more adaptive behaviors in certain contexts (e.g., teach smokers to chew gum in a smoking provoking context). Therapists can influence clients to institute consequences (e.g., if smokers smoke, they might punish themselves, or if they avoid smoking, they might reward themselves).

In spite of the apparent effectiveness of these approaches in the treatment of, for example, fear-related behavior problems, obesity, and addiction-related problems, this classic behaviorist approach has been criticized, as it appears difficult to sustain the view that thoughts and cognitions have no influence on maladaptive behavior. Furthermore, behavioral therapy has been criticized for not finding a good strategy for treating, for example, depression-related pathology.

### Wave 2: Cognitive Behavioral Therapy

Following this dissatisfaction and in line with new empirical findings in cognitive psychology, cognitive behavioral therapy was developed. Aaron Beck and Albert Ellis put the therapeutic emphasis (back) on cognitive processes within the individual, beyond observable behavior. The original S-R chain was expanded and became an S-O-R chain in which the O stands for the organism, reflecting the processes that explain why stimuli in a context lead to certain behaviors. First attempts to explain the O were then made pointing at the role of self-regulation, an ability people can use to control unfavorable behavior (e.g., impulsivity). This led to several treatment programs, including some for children (e.g., children with attention deficit hyperactivity disorder), using new techniques such as self-observation (based on diaries), goal setting, and self-evaluation. Next, it was also assumed that cognitions affect both emotions and behavior, and that therapy should first change the cognitions and emotions, leading consequently to more adaptive behavior. This has led to the well-evaluated stress-inoculation training (learning to regulate emotions and stress through self-instruction for people with psychosomatic problems), incorporating techniques like cognitive restructuring, problem solving, and relaxation training.

Whereas these cognitive processes in earlier times tended to be described very vaguely, Beck developed a theory in which he stated that the development, maintenance, and relapse of depression can be explained by the influence of what he called cognitive schemas. Cognitive schemas were described as “a structure for screening, coding and evaluating impinging stimuli” (Beck, 1964, p. 564). It was assumed that information processing within the organism is more or less a symbolic behavioral chain that is affected by external or internal stimuli (input) and that the outcome is observable in the behavior that one is demonstrating, guided by these cognitive processes (output). On the basis of cognitive schemas, an individual interprets new experiences in a meaningful way. Schemas direct attention to previously stored information to generate expectancies and interpretations for new experiences. Research has shown that depressive individuals have cognitive schemas with less adaptive content, leading to three different biases in the processing of information:

1. A bias in the attentional information processing: If a depressive schema is active, individuals will more easily direct their attention toward negative information.

2. A bias in the interpretation of (ambiguous) information: If a depressive schema is active, it will cause individuals to interpret information more negatively.

3. A bias in the recall of schema-congruent information: If active, a depressive schema will improve the recall of negative information.

These insights have led to the development of several different techniques to challenge and change less adaptive cognitions using, for example, the technique of the Socratic dialogue. This technique has been designed to challenge negative cognitions, not by persuading the client but rather by inviting the client to explore the negative thought, leading the client to question the value of the maladaptive cognition. Besides its use in the treatment of depression, this technique has turned out to be useful in the treatment of social anxiety, bulimia nervosa, panic disorder, trauma, and obsessive-compulsive disorder. Alternative therapeutic strategies for changing maladaptive cognitive processes involve increasing clients’ problem-solving skills, guiding clients through behavioral experiments (during which clients try to do things that challenge their assumptions to see
whether the assumptions are correct), or using psycho-education to provide accurate information about the content of maladaptive schemas.

In addition, Beck’s theory has been further refined and has led to the development of a more encompassing theory with regard to maladaptive schemas that are constructed early in life in interaction with parents or other important caregivers and that later influence the development of psychopathology. In line with the acknowledgment of maladaptive cognitive schemas, Jeffrey Young has developed the cognitive schema theory, distinguishing different so-called early maladaptive schemas and providing therapeutic strategies to tackle the different maladaptive schemas underlying a wide variety of symptoms of psychopathology, including those that characterize axis II disorders (more specifically, personality disorders).

Although the “cognitive revolution” within behavioral therapy has led to the development of often well-established evidence-based cognitive-behavioral treatment protocols, many controversial issues remain. This is partly due to the fact that research on cognitive psychology and cognitive therapy has been developing rather independently, leading to a gap between research findings and therapeutic theory. It also remains uncertain whether focusing on maladaptive cognitions has an additional effect above mere activating alternative behaviors. In addition, incorporating cognitive treatment techniques within behavioral therapy has led to a mix of treatment ingredients that makes it hard to identify which are the real mechanisms of change. Furthermore, there seems to be some slowing down in the development of cognitive therapy with only a few true innovations to improve efficacy. It is interesting that therapists tend to be rather reluctant to use cognitive techniques; this may be because these techniques focus so strongly on rationality and cognitive skills, seemingly at the expense of emotions.

**Wave 3: Acceptance and Commitment Therapy and Mindfulness Based Cognitive Therapy**

Following this dissatisfaction, cognitive behavioral therapists started developing entirely new therapeutic approaches of which acceptance and commitment therapy (ACT) and mindfulness based cognitive therapy (MBCT) currently are the most cited treatment programs. As noted by Steven Hayes and colleagues, these programs have in common that “rather than focusing on changing psychological events directly, these interventions seek to change the function of those events and the individuals’ relationship to them through strategies such as mindfulness, acceptance, or cognitive defusion [a form of cognitive distancing]” (Hayes, Luoma, Bond, Masuda, & Lillis, 2006, p. 4).

The most important treatment strategy according to these new approaches is to help clients experience their internal struggle with unfavorable cognitions or emotions instead of experientially avoiding internal cognitive or emotional information. If the urge to avoid negative experiences guides behavior, individuals will no longer try to accomplish personal goals and pursue personal values. Instead, they will be overly focused on (mostly maladaptive) behavior that aims to avoid negative emotions and interferes with personal goals and values. By receiving training in so-called mindfulness skills, clients learn to stop avoiding negative cognitions and emotions and begin to accept them and to “stay with their mind in the present.” During these training sessions, clients learn to meditate. Clients learn to focus their attention on the actual present and to stop being distracted by (maladaptive) thoughts or emotions. Once they have acquired this ability, clients will be more skilled in accepting their internal struggles instead of following their urge to solve or avoid their emotional distress; in the meantime, they can continue working toward what they consider to be important goals in their lives.

Increasing clients’ mindfulness skills results in a significantly improved quality of life in patients with chronic pain and in clients with recurrent depression. Current research is investigating the effectiveness of ACT and MBCT in treating a very wide variety of symptoms. Although this implies that the value of these new approaches requires more evaluation, the therapeutic community is becoming increasingly enthusiastic about using these new strategies.

In summary, this brief overview shows how behavioral therapy is a therapeutic approach that continues to be “under construction.” Following new findings and developments in experimental, neurological, clinical, and other fields of psychology, the therapeutic strategies continue to be adapted and reevaluated. This evidence-based therapeutic approach is probably one of the most important
values of behavioral therapy, as it improves the quality of mental health care offered. The downside is the often wide gap between research and its translation into helpful techniques that can be successfully applied.

Guy Bosmans and Caroline Braet

See also Anxiety Disorders; Behaviorism; Borderline Personality Disorder

Further Readings


Behaviorism

Just as there are many “cognitive psychologies,” there are many varieties of behaviorism. All behaviorisms hold that behavior (activity) is the subject matter of psychology, in contrast with psychology defined solely as the study of conscious experience. Some forms of behaviorism hold that behavior is significant in its own right. Other forms are actually cognitive (mediational) and treat behaviors as mere indicators of underlying processes. Most current applications of psychology are based on some form of behavioral approach.

This entry constitutes a brief overview of behaviorism through a summary of the views of four well-known behaviorists. John B. Watson founded the movement in 1913 and it was extended and further popularized by Edwin Guthrie in the mid-20th century. During the same period, Clark Hull attempted to transform all of psychology into a rigorous science that still exists as theoretical behaviorism. Finally, B. F. Skinner created radical behaviorism as a philosophy of the science of psychology, but he is better known for the application of behavioral methods in education, the treatment of psychological dysfunction, and elsewhere.

John B. Watson

Although Edward Thorndike introduced behaviorism in 1898, when he showed how the use of rewards and punishers (the law of effect) is far more useful than had been suspected, Watson popularized it, beginning with a combative paper published in 1913 and titled “Psychology From the Standpoint of the Behaviorist.” In that paper, Watson expanded the domain of psychology, which, aside from Thorndike’s work, had been largely restricted to the analysis of verbal reports of conscious experience by adult human subjects. If psychology is the study of behavior, then it need not involve verbal report, and a variety of populations—including children, people with certain mental disorders, and even animals—could be studied.

Mental Activity and the Goal of Psychology

Watson did not “throw out the mind,” as is commonly assumed. He merely followed Thorndike in treating mental activity as a form of behavior. Thinking, feeling, remembering, imagining, and seeing are not causes for what we say and do—they are part of our behavior. He defined the goal of psychology to be the study of stimulus-response relations, emphasizing that stimulus could be a discrete event, like a light flash or a phobic stimulus, or a molar event, extending over long periods, like a method of teaching reading or the redistribution of wealth. Response was treated the same way: Depending on your purpose, it could be an eyeblink or a decrease in crime.

Pattern Reactions

Although Watson had a master’s degree in biology, he believed that behavior was important in its own right and that biological underpinnings were only distractions. Yet, he argued that behavior involves
the whole body and we always see pattern reactions, involving the three groups of behavior. Thus, at any moment we are moving, feeling emotion, and communicating, in different degrees (in Watson’s words, manual, visceral, and laryngeal habits). For example, while making an important decision, I may pace the floor, feel my heart rate increase, and talk to myself. Watson believed that “thinking” involves all of these behaviors, as does anger and imagination and everything else that we do.

**Treat Phobias**

Mary Cover Jones assisted Watson in the testing of at least five methods of eliminating phobic reactions in young children. After treating dozens of young patients, they settled on the method of counterconditioning, later popularized by the psychiatrist Joseph Wolpe and now widely used. Briefly, the method consists of pairing the phobic stimulus with a competing stimulus, such as a favored food or muscle relaxation, in such a way as to substitute relaxation for fear.

Watson’s academic career ended after a 1920 scandal and he went into advertising, where he quickly became successful. He was one of the main figures who transformed advertising from a provider of information about products to a vehicle of persuasion, as it has remained.

**Edwin Guthrie and Stereotypy**

Guthrie was a behaviorist who presented the simplest possible theory that stressed the prevalence of stereotypy. That is, in any situation, we do (think, move, feel) as we did the last time we were in that situation. Because the present situation can never be the same as a previous situation, our behavior varies. To change our behavior, we should use any means necessary to produce the behavior we want in the situations we want. For example, Guthrie suggested that if we want to train a dog to come on command, we would not repeat the command, “Come,” if the dog does not come, because that attaches the command to the wrong response. So, we offer food or use a rope to pull the dog to us. If we want to eliminate a behavior, we substitute an incompatible activity—Guthrie suggested eating apples as an alternative to smoking. However, in his own case he found that it was possible to smoke while eating.

Guthrie felt that rewards and punishments were overemphasized in psychology and that they work only when they result in new behaviors that substitute for previous behaviors. His writings are filled with examples illustrating this process and the reader cannot help but notice Guthrie’s constant concern with practical matters.

Space limits prohibit further discussion of Guthrie’s theory, but the interested reader will find countless practical applications to daily life and to education in his 1952 *The Psychology of Learning*.

**Clark Hull: We Are All Machines**

**Drive Theory**

The 1940s and 1950s were dominated by the behaviorism of Hull and his followers, which epitomized stimulus-response (S-R) learning theory. Educated as an engineer, Hull viewed all organisms as machines constantly acting to maintain homeostasis. This means that we constantly work to maintain ideal levels of nutrients in the body, blood pressure, body temperature, and so on, as popularized by Walter Cannon at Harvard (following the French physiologist, Claude Bernard). Thus, we are always in a state of need for something, whether it be food, water, warmth, comfort, or something else. When a need motivates action, it acts as a “drive.” All through our lives, we act to reduce drives: This was a popular view during the 20th century, held by many others, including Sigmund Freud, who published his views on that subject in 1915. For Hull, rewards (reinforcers) always reduce drives or are associated with other things that reduce drives.

Hull’s major work, *Principles of Behavior*, published in 1943, ensured his leadership in experimental psychology. His theory was laid out as a series of postulates, all of which were true or were widely believed to be true. For example, the first postulate stated that stimulation, say on the skin, leaves an aftereffect that persists for a longer or shorter time depending on the strength of stimulation and other factors. The fourth postulate holds that stimuli present when a response is followed by reinforcement (drive reduction) become attached to that response and form S-R habits, constituting all cases of learning.

Hull insisted on clear language rather than “weasel words,” and he insisted on quantifying everything. He developed equations to predict behavior in
all kinds of situations, and, with his many followers, he was successful enough to influence workers in industrial psychology, social psychology, and those attempting to assess need for achievement and other needs during the 1950s and 1960s. Even in the 21st century, Hull’s influence is clear: S-R psychology’s simplicity appeals to many people.

B. F. Skinner: Not S-R Psychology

In 1938, Skinner published *The Behavior of Organisms*, a book that remains a mystery to many readers. That is because Skinner’s “operant” psychology is very different from other behaviorisms and from all cognitive psychologies. In later publications, he introduced radical behaviorism, a philosophy of science and mind that further baffles many readers. The following sketch cannot fully clarify all of this, but his 1974 popular book, *About Behaviorism*, is a good source for interested readers.

**Radical Behaviorism**

Skinner introduced radical behaviorism in 1963, in a piece titled “Behaviorism at Fifty,” a reference to Watson’s 1913 diatribe. He contrasted this with methodological behaviorism, which is the view of the old S-R psychologies and of cognitive psychology. Methodological behaviorists stress scientific method and observable behavior: They believe that we can study the body and other physical things but not the mind, as one’s mental activity cannot be observed by others.

Skinner’s radical behaviorism is different and questions the mind/body distinction. The real distinction is between private/public, and psychology must include both. In subsequent writings, he proposed ways in which we may know the private (mental) world of another. Thus it is incorrect to say that Skinner opposed the study of the mind.

**Theories Are Unnecessary**

How could Skinner argue against the use of theories while he appeared to be introducing one of his own? The answer lies in what he meant by theory, of course. He felt that, as long as we refer to behavior (including thinking, seeing, and the rest) and the environment (stimuli, consequences of responding), no theory is involved, but when we refer to underlying causes of behavior, such as cognitions, S-R habits, brain mechanisms, or information processing mechanisms, we have a theory and usually neglect the behavior that we are ostensibly explaining.

**Time and Contingencies**

Often, theories are concocted because we do not appreciate the fact that behavior occurs over time and cannot be understood if we examine only discrete snapshots. As Aristotle taught, one swallow doesn’t make a summer and one honest act doesn’t constitute honesty. Love, intelligence, and memory cannot be understood except as patterns of behavior occurring over time. To take a simple example, I imagine the house I grew up in, and I can do this because of a long history living there as a child and because of frequent recollections. But we might teach children that we have the “image” of the house in our “mind,” because a child cannot understand the concept of behavioral history. Both William Baum and Howard Rachlin elaborated and extended this point of view.

The emphasis on “behavior over time” led to the development of schedules of reinforcement in the 1950s; for the first time, behaviors could be precisely controlled over days, weeks, months, or years. This approach could explain such diverse phenomena as why gambling casinos are so successful and how to arrange payoffs and penalties to improve education, increase workers’ production, and treat psychological maladjustment. *Contingencies* refer to arrangements of consequences for specific behaviors in specific situations, exemplified in reinforcement schedules and in myriad applications to human problems. In important respects, Skinner’s work represents a return to Watson’s way of thinking, although Skinner never thought so.

**Inspiration**

Skinner’s greatest influence may lie in the inspiration he lent to countless others. The Association for Behavior Analysis International (ABAI) included more than 16,500 members in 2008/2009. He also inspired some of the leading figures in modern psychology, from Baum and Rachlin to the social psychologist Daryl Bem, whose theory of self-perception exemplifies Skinner’s thinking.

**Cognitive Revolution?**

In the 1960s, a so-called cognitive revolution occurred in reaction to popular interpretations of
behaviorism. It was fueled in part by fascination with the newly invented digital computer and by fresh interest in species-specific (instinctive) behavior. However, the appropriate target of the rebels was not the behaviorism of Watson or Skinner, as many critics believed; it was the simplified S-R psychology derived from Hull’s theory. Thus, Skinner’s followers never understood why criticism was directed toward them. Even worse, the rebels were themselves largely S-R psychologists, and their “revolution” was no more than a changing of theoretical terms. For example, S-R habits became parts of an information processing flow chart. Thomas Hardy Leahey has argued that the “cognitive revolution” was a myth.

John C. Malone

See also Behavioral Therapy; Reinforcement Learning, Psychological Perspectives

Further Readings


Belief and Judgment

We believe many things about the world around us. We also make conscious judgments as to how things are. Much philosophical attention has been devoted to belief; judgment has been relatively neglected. This entry sketches central themes in discussions of both, emphasizing their joint importance.

Propositions

Beliefs and judgments are relations to propositions. If John and Jean, respectively, utter the sentences “Paris is extraordinary” and “Paris est extraordinaire,” they both say the same thing with their different words. If Jean and John both believe (or judge) what they say, they both believe (judge) the same thing. In philosophical parlance, they stand in the believing (or judging) relation toward the proposition expressed by their sentences, namely, that Paris is extraordinary. Other mental states and acts exhibit this form. If Amandine desires that her steak is blue, she stands in the desire relation toward the proposition that her steak is blue.

Beliefs and judgments are true or false depending on the truth or falsity of their objects. We talk of believing in ghosts or believing a person. Because ghosts and people are not themselves true or false, philosophers typically assume that all belief talk can be reconstructed employing propositional complements apt for truth-evaluation. If Keith believes in ghosts, he believes that ghosts exist. If Jennifer believes Jude, she believes that Jude is telling the truth.

Controversial issues arise in individuating beliefs and judgments. Here is one problem case from Saul Kripke. Peter fails to realize that Paderewski is both a Polish statesman and a pianist. With the pianist in mind, Peter sincerely states, “Paderewski has musical talent.” With the politician in mind, and believing that all politicians are entirely unmusical, Peter sincerely declares, “Paderewski has no musical talent.” Peter seems to make contradictory claims. Given his sincerity, does this reveal contradictory beliefs? If so, it would show strikingly that no matter how rational and reflective Peter is, he can unwittingly harbor contradictory beliefs and, indeed, consciously entertain contradictory judgments.

Belief Versus Judgment

Judgments are occurrent, conscious events. One cannot judge anything during dreamless sleep. In contrast, beliefs are standing attitudes or enduring states. Beliefs can be retained for years without once crossing one’s mind. Beliefs can be acquired and lost. But these events must not be identified with the beliefs themselves. Judgments often lead to the acquisition (or loss) of a belief. Thus, I may work out the formula for the surface area of a sphere and
conclude that the answer is $4\pi r^2$. This judgment may lead me to acquire the belief that the formula is $4\pi r^2$. However, judgments do not always lead to belief acquisition. In an oral exam, my judgment may simply manifest my long-standing belief that $4\pi r^2$ is the formula. Moreover, already believing that the formula is $4\pi r^2$ because my professor has told me, I can set out to prove it to myself. If successful, that process will end in the judgment that $4\pi r^2$ is the formula, but no new belief is acquired.

Is judgment always the acquisition or manifestation of belief? An example adapted from Christopher Peacocke suggests not. A department head sincerely judges that gender is no indication of applicant quality. However, for years, his hiring behavior has judges that gender is no indication of applicant quality. He seems justified. In 1963, Edmund Gettier offered two famous counterexamples apparently showing that a subject can have a justified true belief that $p$ and yet fail to know that $p$. Subsequent work (beyond the scope of this entry) attempts to provide a better analysis of knowledge in terms of belief, or to show that there can be none.

We can believe falsehoods, but deliberation as to whether to believe that $p$ appears to be guided solely by deliberation as to $p$’s truth. We do not, for example, consider whether we’d like it to be the case that $p$. Controversy surrounds the formulation and explanation of this transparency thesis. One issue is whether truth is the norm of belief, in that one ought to believe $p$ only if $p$ is true. Another is whether we have any epistemic freedom to choose what to believe. Active deliberation is a matter of reaching or suspending judgment. Moreover, judgments (because they are actions as opposed to states) are of the right category to be free or constrained. These issues must be considered in relation to judgment and not just belief.

The relationship between transparency and self-knowledge provides a case in point. Sometimes we discover what we believe in the same way that third parties do. Thus, the department head in the earlier example may acknowledge his sexist beliefs when confronted with the systematic bias in his hiring decisions. However, normally, we take ourselves to enjoy special first-person access to our beliefs. Some suggest that transparency helps explain this. Asked if you believe that $p$, (by transparency) you simply consider whether $p$. If you conclude that $p$ is true, you are justified in believing that you believe that $p$. It is initially opaque how endorsement of a proposition concerning the world could justify a proposition concerning one’s mind. However, once we recognize that judgments are actions, we can exploit our distinctive agential perspective in giving an account of self-knowledge. Recall Soteriou’s view that one judges that $p$ just if one concurrently believes that one is judging that $p$.

Peter Geach suggests that because the whole content of a judgment must be grasped at once, judgments must be instantaneous. This makes it hard to see how judgments could be conscious mental actions as opposed to mere changes of mental state. Soteriou replies that judgments cannot be characterized exhaustively in terms of their contents. A judgment that $p$ always has some unfolding conscious mental event (e.g., the entertaining of speech sounds in one’s auditory imagination) as its vehicle.

Epistemology

Knowledge is a factive mental state: If you know that $p$, $p$ is true. Neither belief nor judgment is factive. Traditionally it was assumed that belief constituted knowledge when true and sufficiently justified. In 1963, Edmund Gettier offered two
it. Impressed by this, some philosophers introduce the idea of a subject’s degree of belief, or credence, in a proposition. A natural way to try to establish a subject’s credence in a proposition is by establishing the shortest odds at which they will accept an affordable bet regarding its truth. The shorter the odds they accept, the greater their credence will be. Thus, Stephanie has a credence of 0.25 in Brazil winning the World Cup just if she will rationally accept a bet at odds longer than 3:1 against Brazil winning.

Some philosophers reject degree-talk entirely, arguing that degrees of belief are better thought of as outright beliefs in objective chances. Accordingly, Stephanie does not have a credence of 0.25; rather, she believes outright that there is a 0.25 chance of Brazil winning. Note that because we cannot know, based on the odds, that we will not win a very large lottery, it is implausible to equate knowledge with sufficiently high credence.

Analyzing Belief

Many philosophers have sought to analyze belief as a kind of disposition. We can manifest our beliefs in actions (and reactions) other than judgments. Thus, belief cannot simply be the disposition to judge that \( p \). More plausibly, to believe that \( p \) is to be disposed to act as if \( p \) were true. Thus, the (repressed) sexist belief of the department head is manifested in his overall pattern of sexist actions. A given belief explains an action only in concert with other beliefs and desires. I manifest my belief that the left-hand track is the quickest way to Tipperary by taking it, only if I want to go to Tipperary quickly. Even if I do have that desire, I may believe that my nemesis is lurking to ambush me on that path and so take another. Or again, I may take that path precisely because I want a final showdown. There is no prospect of spelling out what it is to act as if \( p \) is true independently of other beliefs and desires. Given the right complement of beliefs and desires, any action can be thought of as manifesting a given belief. Consequently, philosophers argue that we must treat propositional attitudes holistically, not one by one.

Dispositional analyses emphasize what belief leads to. We might also consider what leads us to belief. According to functionalist approaches, a subject has a belief that \( p \) just if they are in a state that occupies a certain causal role specified in terms of inputs (e.g., seeing flames) and outputs (e.g., running away) as well as relations to other mental states. Judgments will typically form a crucial class of these inputs and outputs. Functionalism is an extremely broad church. Its members divide as to whether psychological roles are determined by conceptual analysis or scientific investigation. They also divide as to how inputs and outputs should be specified. Those hoping to reduce the mental to the physical attempt to spell out inputs and outputs in purely physical terms. Others argue that functional specifications cannot be provided except in mentalistic terms.

One key disagreement concerns the importance of a subject’s internal (physical) makeup in belief attribution. According to the interpretationist approach associated with Donald Davidson and Daniel Dennett, a subject is held to believe that \( p \) just if (roughly speaking) the entirety of their behavior is best explained by a set of beliefs, desires, and other attitudes that includes the belief that \( p \). Internal structure is strictly irrelevant. Interpretationists embrace the apparent consequence that whether a subject believes that \( p \) may be indeterminate, because two different but equally explanatory theories of behavior may be available. Representationalists, in contrast, take beliefs to be causally efficacious internal representations, realized in the neural architecture of the brain. One has an attitude toward a proposition \( p \) by being related to a stored representation with \( p \) as its content. What makes this attitude a belief as opposed to a desire is the functional role the representation plays in a subject’s mental economy.

Representationalists diverge when it comes to specifying the nature of the stored representations and explaining how they acquire content. Fodor, in his influential language of thought hypothesis (designed to explain the human capacity to entertain a potentially unlimited number of different thoughts), claims that our internal representations have a structure like that of natural language: Complex representations are built out of subsentential symbols, the content of a complex representation being a function of the content of the atomic symbols it contains in combination with the language’s syntax. On this picture, thought, in general, involves the recombination of symbols in a way that is causally sensitive to syntactic rules. There remains a great deal of disagreement concerning the status and commitments of this view. One central issue is
how it fares with respect to connectionist models of cognitive architecture.

Ian B. Phillips

See also Attitudes and Behavior; Conscious Thinking; Dutch Book Arguments; Representational Theory of Mind

Further Readings


Bilingual Language Processing

The increasing global prevalence of bilingualism has led to a growing interest in how bilinguals process language. Research on bilingual language processing focuses on bilingual comprehension and production skills. The study of these topics has helped researchers answer a variety of questions, including the following: Can bilinguals ever “turn off” one of their languages? How are words and translations interconnected in bilingual memory? What word and learner characteristics influence the way words are represented and processed? To address these questions, researchers employ a number of tasks, some of which are discussed in this entry. Also discussed is the nature of bilingual memory representation and how it is influenced by learner and word characteristics. Models of bilingual memory representation are considered in the context of these issues.

Tasks

A number of tasks are used to examine bilingual language processing. In some of these tasks, the speed and accuracy with which people respond are used to infer how words are represented in bilingual memory and/or how these representations are accessed. In the “translation production” task, a word is presented in one language to be translated aloud into another language. “Translation recognition” is a similar task, wherein a pair of words is presented to be judged as correct or incorrect translations. In the “picture naming” task, a line drawing is presented to be named in a given language. In the “lexical decision” task, a letter string is presented to be judged a real word or a nonword in a given language. Finally, in the “priming” task, a *prime* is shown (typically very briefly), followed by a *target*, to which an individual must make a response (usually, a lexical decision). Using words of various types in these tasks allows researchers to infer how these kinds of words are represented and accessed.

In some cases, these tasks are performed following a sentence context, often to determine the way context affects how words are processed. Other tasks specifically examine sentence-level processing. The “self-paced reading” task involves presentation of a sentence one word at a time, paced by the bilingual. The “grammaticality judgment” task involves presenting a sentence for a judgment about grammatical acceptability. Both of these tasks allow researchers to examine bilinguals’ sensitivity to grammar in their two languages, often by including violations of grammar.

Selective/Nonselective Access

A prominent theme in bilingual language processing research is whether bilinguals are able to “turn off” one of their languages when they are using the other exclusively (i.e., can they selectively access one language?). This issue has been explored using several tasks that provide converging evidence; some of these tasks exploit the relationship between words in the two languages to determine whether both languages are accessed. Although initially the question of interest was whether bilinguals typically access
only one language at a time, current research aims to understand when such selective access is likely to occur and what circumstances increase its likelihood. Some research on this issue suggests that bilinguals can more easily “turn off” a weaker language than a very dominant language (e.g., their first language).

Second Language Proficiency
A learner characteristic that has been explored in relation to language processing is the proficiency of the bilingual. The revised hierarchical model by Judith Kroll and Erika Stewart provides a basis for understanding how increased proficiency in a second language would lead to changes in how second language words are connected to their meanings and, in turn, to changes in the way that these words are processed. This model proposes that initially, second language words only gain access to their meanings through the first language, but that when the bilingual becomes more proficient in the second language, the connections between second language words and their meanings become stronger. By contrast, the distributed feature model by Annette De Groot proposes that learners have direct access to the meanings of second language words from the beginning stages of learning, and that word characteristics (described in the next section) play a more important role in lexical processing.

Word Characteristics
In some cases, researchers focus on the way that certain types of words influence bilingual language processing. Two specific characteristics have been studied in detail: form overlap and word concreteness.

Form Overlap
One characteristic of cross-language word pairs that influences bilingual language processing is the relative similarity of the words in spelling, sound, or both. Translation equivalents (i.e., words that mean the same thing across languages) that are similar in form are called cognates (e.g., teléfono in Spanish and telephone in English). In many tasks (e.g., translation production, picture naming), cognates elicit faster and more accurate processing. By contrast, cross-language word pairs that share form but not meaning are referred to as interlingual homographs or false cognates (e.g., pan in Spanish and bread in English). This type of cross-language relationship often slows processing or makes it less accurate (e.g., in translation production). The distributed feature model hypothesizes that cognate translations share more of their meaning features than do non-cognate translations across languages.

Word Concreteness
Word concreteness has been studied extensively with respect to its influence on both monolingual and bilingual processing. Concrete words refer to perceptible entities whereas abstract words refer to imperceptible entities. In many tasks (e.g., translation production), concrete words are processed more quickly and accurately than abstract words. As with cognate translations, the distributed feature model proposes that concrete words share more meaning features across languages than do abstract words, leading to the concrete-word advantage in bilingual translation. Recent research has revealed some conditions under which the concrete-word advantage is not observed, both within and across languages.

Natasha Tokowicz

See also Bilingualism, Cognitive Benefits of; Concepts and Language; Heritage Language and Second Language Learning; Language Development; Production of Language

Further Readings
relationship between language and other cognitive abilities. Interest in the topic has extended beyond the academic literature into the popular press because of reports that bilingualism improves some aspects of cognitive functioning. This entry summarizes some of the more recent literature on how bilingualism affects language and cognitive processing. Evidence suggests some significant differences between bilinguals and monolinguals that, although small, have interesting implications for theories of language processing and cognitive control. The effects reviewed pale in comparison with more obvious differences associated with bilingualism (e.g., bilinguals can communicate with a broader audience). However, subtle differences between bilinguals and monolinguals have contributed to important theoretical debates in the literature, and increasingly bilingualism is viewed as a tool that reveals the fundamental mechanisms supporting language and other cognitive processing. A current challenge in this line of work is to link the benefits of bilingualism more precisely to their cause.

**Bilingual Processing Advantages**

Evidence is accumulating that bilinguals outperform monolinguals on a variety of tasks. These benefits of bilingualism emerge in infancy and persist throughout the life span into old age. Beginning with the pioneering work of Ellen Bialystok, initial findings were that bilingual children are better at “meta-linguistic” skills. For example, bilingual children separated meaning and grammar at an earlier age than did monolinguals when asked to judge if sentences like “Why is the cat barking so loudly?” are grammatical or not. Producing the correct response (“yes” in this case) requires identifying that words and their referents are separate entities and focusing on grammar while suppressing the tendency to focus on meaning (i.e., in this case that cats don’t bark). In several additional studies, Bialystok and colleagues reported that bilingual children are advantaged in tasks that seem to have nothing to do with language. For example, another study found that Cantonese-English bilingual children were better than monolingual children at switching card-sorting rules (e.g., sort by color vs. sort by shape). Similar advantages are found in young adulthood, although these tend to be smaller and more difficult to observe than advantages seen in children and in aging studies. In a series of studies by Albert Costa and colleagues, young adult Spanish-Catalan bilinguals were faster than monolinguals to indicate the direction of a center arrow in a task with congruent (→→→→) and incongruent (→→→→→) flankers. In another study by Anat Prior and colleagues, college-aged immigrant bilinguals were faster than monolinguals to switch from classifying objects by color versus by shape. With aging, bilingual advantages seem to become more robust. One study by Bialystok and colleagues that generated considerable press found that bilingualism in immigrants to Canada was associated with later age-of-onset of Alzheimer’s disease (AD) relative to that typical for monolinguals. Here it is hypothesized that bilingualism may be analogous to **cognitive reserve**, which also delays the onset of AD; for example, high levels of education may delay the onset of AD.

**Cognitive Mechanisms Underlying Bilingual Advantages**

An important goal in research on bilingualism and cognitive control is to identify the connection between theories of bilingual language processing and the observed advantages. To date, the precise link between theory and the advantages found remains elusive and is the focus of some debate. On one view, bilinguals’ practice with resolving competition between languages leads them to develop more efficient cognitive mechanisms for resolving competition. This hypothesis relates to David Green’s inhibitory control model, which proposes that bilinguals face constant competition between languages. However, attempts to link bilingual advantages specifically to inhibitory control (the ability to inhibit responses or information that comes to mind automatically) have had mixed results. Bialystok and colleagues suggest that bilingual advantages are restricted to complex tasks that require control over attention to competing cues (interference suppression), but are not present in tasks that trigger competing responses (response inhibition). In a similar vein, work by Albert Costa, Mireia Hernández, and their colleagues in Barcelona, where bilingualism is extremely common, suggests that bilingual advantages may reflect heightened monitoring ability (e.g., in Barcelona, it is common for children to
speak one language with one parent and a different language with the other parent—an arrangement that requires constant language switching and monitoring).

**Apparent Failures to Replicate**

Important clues as to the mechanism underlying bilingual advantages may emerge from cases in which attempts to find bilingual advantages are not successful. Costa and colleagues have described bilingual advantages as being difficult to replicate or “now you see them, now you don’t.” An important consideration is to differentiate studies that may simply have lacked power for observing bilingual effects from those that can provide an elegant demonstration of how bilingual effects can be made to appear or disappear with only subtle changes in the task. In other cases, no bilingual advantage may be found even though a positive influence of bilingualism is present if other factors correlated with bilingualism negatively affect executive control (e.g., bilingualism is associated with low socioeconomic status in some parts of the world).

The difficulty of observing bilingual effects across bilinguals of different types also suggests multiple mechanisms underlying the observed advantages; for example, some types of bilingualism may lead to a switching advantage (i.e., faster responses when switching from one task to another), whereas others may lead to better ability to monitor which response is appropriate when. An important clue as to the origin of bilingual advantages is that “more bilingual” sometimes seems to translate into “more advantage.” Here “more bilingual” can mean several different things, including knowledge of three languages (tri-lingualism), knowledge of more than three languages (multilingualism), greater degrees of proficiency (native-like ability in both languages), or degree of language use (equal use of both languages in daily life). In one such study, Andrea Mechelli and colleagues found that bilinguals had denser gray matter in left inferior parietal cortex when compared with monolinguals, with increasingly greater density for bilinguals with earlier age-of-acquisition of both languages and for bilinguals with greater degrees of proficiency in a second language. Although suggestive, this study could not reveal precisely what it is about bilingualism that leads to the advantages observed because language proficiency was measured using an extensive battery of tests (and correlations for individual subtests were not reported). A challenge with studying “better bilinguals” is that such people may have started out with a stronger set of cognitive skills that allowed them to become highly proficient bilingual or multilingual (the classic “chicken or egg” problem: Does multilingualism lead to advantages, or are people who are advantaged more likely to become multilingual?). Some degree of connection between bilingualism and general mechanisms of cognitive control comes from imaging studies showing overlapping brain regions active during task switching and language switching; however, current technology may be insensitive to subtle differences between task-general and language-specific control.

**Cognitive Costs Associated With Bilingualism**

On the flip side of bilingual advantages, there is some clear burden associated with learning and maintaining more than one language in a single cognitive system. Work by James Flege and colleagues indicates that bilinguals (even highly proficient early learners of two languages) perceive and produce vowels slightly differently from monolinguals, probably reflecting cross-language interference. Studies by Kimbrough Oller and colleagues have shown that bilingual children tend to know labels for some concepts in one language but not in the other language and vice versa, leading bilinguals to obtain lower vocabulary scores than monolinguals on standardized tests. It used to be thought that bilingual kids “catch up” to their monolingual peers as their experience with both languages increases. However, Tamar Gollan and her colleagues have shown that with sufficiently difficult tests, the vocabulary and naming disadvantages associated with bilingualism can be detected into adulthood and older age. Young adult and older bilinguals are slower than monolinguals to name pictures, have reduced verbal fluency, and are more likely to fail to retrieve words they are sure they know (i.e., have more TOTs or tip-of-the-tongue states). These disadvantages appear even when bilinguals are tested in the language they report is their stronger language.
It might seem elegant to assume that both bilingual advantages and disadvantages are the result of competition between language systems, but Gollan and colleagues suggest that it is a mistake to attribute all of the consequences of bilingualism for language and cognitive processing to one mechanism. Unlike bilingual advantages, which seem to become more robust in older age, bilingual disadvantages may diminish: There is some evidence that older bilinguals do a better job than younger bilinguals in acquiring and maintaining fluency in two languages. Because cognitive control is known to decline with older age, these data also imply that the link between cognitive control and bilingual language proficiency is relatively weak. Instead, bilinguals may effectively be less proficient because they use words in each language less frequently than monolinguals, who are, in a sense, “over-practiced” in the one language that they know. From this perspective, older bilinguals may have an easier time than younger bilinguals in being bilingual, because their increased age allows them more time to master two languages.

Future Directions in Research on Bilingualism

Future work on the effects of bilingualism will be focused on determining which specific aspects of bilingualism are associated with which advantages (and disadvantages) and on how these effects relate to theories of bilingual language processing and to larger conclusions about relationships between language and cognitive control. Although the effects of bilingualism are often subtle, significant effects hold promise for revealing properties fundamental to language processing and to attention/control systems. Indeed, from some theoretical perspectives, it is surprising to find anything more than a very general relationship between language processing skills and performance on nonlinguistic tasks. In this respect, the field seems to have moved far beyond a simple “Is bilingualism good for you or bad for you?” approach to a more principled attempt toward examining the cognitive advantages and disadvantages of bilingualism and their implications for understanding the minds of bilinguals and monolinguals alike.

Tamar H. Gollan

See also Aging, Memory, and Information Processing Speed; Bilingual Language Processing; Heritage

Further Readings


Black English Vernacular (Ebonics)

This entry briefly describes the history of Black English Vernacular (BEV) with primary emphasis on the United States; BEV is frequently referred to as Ebonics, a term that is internationally relevant to the linguistic consequences of the African slave trade. Some changing linguistic terminology related to the speech of enslaved Africans is then introduced, followed by specific linguistic examples of BEV. Concluding remarks confirm the long-standing vibrancy of BEV, or Ebonics, and its contemporary influence on youth culture and youth language usage throughout the world.

Historical Considerations

Africans who were enslaved hundreds of years ago were not able to preserve their native languages. They were forced, on capture and during the ensuing Atlantic crossing, to listen attentively to their captors who, whenever possible, separated slaves who spoke the same language in the hope of preventing insurrections. The initial linguistic contact between slaves and their European captors produced several contact vernaculars, commonly identified as pidgins. Pidgins are distinctive because they have no native speakers. The children of African slaves who spoke pidgin were the first to convert these new contact vernaculars into various Creole languages; this occurred wherever African slaves were sold. In Brazil, the dominant European contact language was
Portuguese. In Haiti, the dominant European contact language was French, and in Jamaica and the United States, the dominant European contact language was English. Ebonics, first coined by Robert Williams in 1975, was originally formulated as a term to be applied broadly to the entire linguistic consequences of African slavery since its inception.

Changing Linguistic Terminology
Prior to Williams’s formulation of Ebonics, linguists used various terms to refer to Black speech. In 1933, Leonard Bloomfield described the difference between substandard English, which for many was considered improper, and nonstandard English, which linguists had identified as being different from prescribed standard English but not inferior to it. In an effort to promote greater accuracy, nonstandard Negro English was eventually adopted by linguists during the 1960s to refer to the speech of American slave descendants; in 1972, the terms Black English and Black English vernacular prevailed. Early in the 1980s, some linguists began to adopt the term African American vernacular English in an effort not merely to provide a racial reference but one that was more specifically tied to the history and culture of enslaved Africans in America as well as their descendants. Ebonics, although coined in 1975, did not frequently appear in technical linguistic studies until after it had gained global attention in 1996 at the behest of the Oakland, California, school board after their passage of a controversial resolution declaring Ebonics, not English, to be the official language of the 28,000 African American students who were enrolled in the school district at that time.

Illustrative Examples of Black English Vernacular
BEV differs from standard American English in several ways, most notably including grammatical differences and phonological variability.

Grammatical Examples

Habitual use of be
They be telling stories (i.e., They are always telling stories).
He be laughing (i.e., He is always laughing).

Copula absence: missing is or are
He coming home tomorrow (i.e., He is coming home tomorrow).
We tired (i.e., We are tired).

Completed actions with be done
She be done finished work early (i.e., She has already finished work early).
They be done told her to go (i.e., They already told her to go).

Use of been with heavy stress on the word itself
I been know your name (i.e., I have known your name for quite a while).
She been sold that car (i.e., She sold that car long ago).

Phonological (Pronunciation) Examples

Final /f/ for /θ/
Ruth → /Ruf/, both → /bof/, tooth → /toof/

Postvocalic /r/ absence
Door → /Doe/, Four → /foe/, Car → /Ca/

Loss of final stop consonant in consonant clusters
Last → /las/, Desk → /des/, Told → /tol/

Phonological metathesis
ask → aks

Contemporary Relevance of BEV and Ebonics
Although the speech of African Americans today is quite diverse—spanning the entire spectrum from BEV to impeccable renditions of the most eloquent forms of standard American English—youth culture, and hip-hop culture in particular, it has been greatly influenced by BEV. Urban poets, rap artists, and hip-hop devotees routinely judge the fluency and artistry of contemporary BEV during highly stylized verbal jousts that are commonly called Battles. Despite the fact that BEV and Ebonics are the historical product of racially sanctioned linguistic isolation, resulting from overt racial segregation and inequality, the
popularity of the Blues, gospel music, and other African American verbal art forms now have cultural appeal that far exceeds people of African descent.

BEV and Ebonics have come to influence the language and culture of many adolescents and young adults throughout the world, particularly if they identify positively with hip-hop culture and genres of speaking that owe their very existence and inspiration to the linguistic legacy of the African slave trade and the long-standing survival of BEV in the United States.

John Baugh

See also Heritage Language and Second Language Learning; Language Development; Word Learning

Further Readings


BLINDSIGHT

Blindsight, as the name implies, is a paradox—a sort of seeing without “seeing.” Ever since the term was first coined in the early 1970s it has raised profound and intriguing questions about what it means to “see,” and what it means to “guess.” For example, if you make 100 random guesses in a row and 90 of your guesses are correct, does it qualify as “guessing”? If you respond accurately (by guessing) to visual stimuli but deny having consciously seen anything, does this count as “seeing”? Barring a paranormal explanation, how could this be possible? By raising these and other questions, the phenomenon of blindsight has enriched our understanding of consciousness, vision, and the brain, and has been fodder for a rich scientific and philosophical debate. This entry will cover the essential concepts from neurology that are relevant to blindsight, provide a brief history of blindsight research, describe the defining characteristics of blindsight and how these might be explained by the architecture of the brain, and discuss the possibility of blindsight in normally sighted individuals as an emerging direction in blindsight research.

Hemianopia (“Cortical Blindness”)

To clearly understand what blindsight is, it is necessary to understand cortical blindness. Lesions in the primary visual cortex (called V1) of the occipital lobe, due to stroke or head trauma, normally lead to blindness in the corresponding region of the patient’s field of view (FOV). For example, damage to the lower part of V1 in the left hemisphere will produce a “blind region” (or scotoma) in the upper-right quadrant of the FOV. Damage to the entirety of V1 in the left hemisphere will lead to blindness in the entire right-hand side of the FOV (called hemianopia, meaning “half-field blindness”). Note that in V1 the signals from the two eyes are combined into a single binocular view (you do not normally see two of everything, even though you have two eyes), so the blind area in this case is not specific to one eye or the other. The two eyes may continue to function normally, but the signals from the left side of each retina (in the latter example) are cut off at their main junction with the cerebral cortex (V1).

Just what do such patients experience when they look at the visual scene? One might suppose that these patients experience something like a large black stain in their visual field that occludes objects from view. In fact, what the patients experience in their scotoma is probably more like what you “see” behind your head. Most likely you do not feel that anything is missing from your field of view behind your head—it is simply not part of your field of view. Likewise these patients are usually unaware that anything is missing, even when they bump into something.

Normally the extent of such patients’ scotomas is determined by asking the patients to fix their gaze and then recording their responses to brief flashes of light, presented one at a time at a broad array of points covering the normal extent of the visual
field (a process called \textit{perimetry}). The points that the patients consistently fail to detect are taken to delimit the \textit{scotoma(e)}, and thus concludes the test—or so it would seem.

**The Discovery of Blindsight**

Toward the end of World War I, the British neurologist George Riddoch examined soldiers who had suffered gunshot wounds to the back of the head (with damage to the primary visual area, V1). He noted that these patients often consciously detected motion within the blind region(s) of their field of view, although it was, and remains, unclear that they actually “saw” anything in the strict sense. Because Riddoch did not offer an explanation, his observations were not followed up.

Then, in the early 1970s, some intrepid researchers had the novel and counterintuitive idea of asking a patient to make guesses about visual stimuli, presented within the scotoma, \textit{even when the patient denied sensing anything at all}. The patient found this to be an unusual request, but complied nonetheless and made “random” guesses regarding, for example, the position of a circular patch of light or the orientation of a small line (horizontal vs. vertical). The patient’s guesses were impressively accurate—much more so than would be expected by chance (e.g., 29/30 correct on one task). In a case study published in 1974, blindsight researcher Larry Weiskrantz and his colleagues reported that, when told how well he had done, the patient “expressed great surprise, and reiterated that he was only guessing” (p. 712). Thus, the term \textit{blindsight} was coined to describe the ability to make accurate guesses regarding stimuli that were not consciously seen. At least several dozen such patients have been identified and tested (quite extensively in several instances) since the time of these early studies. Although blindsight was originally discovered in the context of cortical blindness, the possibility of “seeing without seeing” might also apply to normally sighted persons (see the final section to learn more).

**Controversies in Blindsight Research**

The concept of blindsight is simple, even if counterintuitive and difficult to grasp at first, but the reality is not so simple, and blindsight research (ongoing since the 1970s) has not been without controversy. Early critics suggested that blindsight might simply be an effect of scattered light making its way into the normally sighted region(s) of the visual field or might be mediated by “islands” of spared tissue in the primary visual cortex. Perhaps of greater significance was the suggestion that the nonconscious aspect of blindsight was “trivial,” being no different from normal near-threshold vision (i.e., normal vision when stimuli are so faint as to be practically invisible). At the visual threshold, it is not unusual for normal subjects to show some sensitivity, when \textit{forced-choice} procedures are used, even though they may not be certain of having seen much of anything. Each of these challenges has since been countered, by argument, appropriate control experiments, and/or neuroimaging, and the consensus at present is that no combination of these factors can adequately explain blindsight.

Another area of controversy concerns the unusual conscious sensations that blindsight patients often report (very similar to Riddoch’s patients). Does this count as “seeing”? When a particularly salient stimulus is presented—for example, a high-contrast pattern of lines (a grating) that moves abruptly or flickers—patients sometimes report a sensation but find it difficult to describe and seem reluctant to call it visual. A particularly well studied hemianope, GY, who has been the subject of many studies, sometimes reports a feeling “that something happened” when a high-contrast stimulus abruptly appears in his blind hemifield, but he insists that he does not “see” anything. He has also expressed it as something like “black on black.” Blindsight researcher Petra Stoerig once cleverly asked GY to vary an image presented to his sighted hemifield until it gave an impression of how he experienced an image simultaneously flickering in his blind hemifield. GY complied with these instructions and produced a cloudy-looking dark bar with blurred edges, on a cloudy-looking dark background. This was cited as evidence that GY’s blind-field sensations may indeed be “visual” in some sense but is still inconclusive. Most people are capable of describing a nonvisual sensation in visual terms if asked to do so—like making a visual rendering of a texture or shape that is examined only by touch.

**True Blindsight: The Philosophical Zombie and the Importance of “When”**

Still, if visual stimuli are stationary with moderate to low contrast, the patient may deny experiencing
anything whatsoever and yet continue to make highly accurate guesses. This “pure” form of blindsight has attracted a great deal of attention from philosophers who often cite blindsight in relation to the concept of a “zombie”—an imaginary being that, in its appearance and behavior, is indistinguishable from a human being but that has no conscious experience. Note that the word zombie is used in a specific way by philosophers as part of a particular thought experiment and is not to be confused with the zombies in horror movies or voodoo culture. For a zombie there are stimuli (such as “Hello, how are you?”) and responses (“I’m fine. How are you?”), but otherwise “it’s all dark inside.” The canonical blindsight patient is often cited as something of a “visual zombie,” suggesting that the patient can perform any kind of visually informed behavior just by guessing. Were this true, it would certainly be of clinical value, as it could help to restore visual function in the patient. Unfortunately, for such patients, however, this is far from the truth. In fact, it is probably safe to say that blindsight, although of enormous scientific value, seems to be of no spontaneous practical use to the patient (although valiant efforts are under way to train patients to use it).

One of the main reasons for this, and one of the most intriguing and overlooked facts about blindsight, is that the patient has to be told when to guess. For example, they may be told, “When you hear the beep, guess the orientation of the line on the computer screen.” If no conscious cue is given to the patient as to when to respond, and the stimulus provokes no conscious sensation whatsoever in the patient, then no response is made—the patient simply waits. This observation brings to light what might be a defining feature of conscious (as opposed to nonconscious) sensory information processing in the brain: Although the brain can apparently discriminate between two stimuli (e.g., a horizontal vs. a vertical line) without consciously seeing them, detecting a stimulus (e.g., “raise your hand when you see the light flash”) seems to be intimately tied to having a conscious experience of some sort. The philosopher Daniel Dennett, in his book Consciousness Explained, went so far as to predict that, if a blindsight patient could somehow be trained to guess when to respond, the ability to detect a visual event would necessarily bring with it a conscious sensation of some sort. This prediction has yet to be formally tested.

The Limits of Blindsight

Even when the patient is given a cue to respond, the range of visual properties that a blindsight patient can accurately discriminate by guessing (or by indirect measures such as the priming of reaction time) is limited. Among the stimulus properties shown to be discriminable via blindsight are presence or absence of a stimulus, motion, flicker, location of a stimulus, and line orientation. There is also evidence for sensitivity to the emotional content of a stimulus (e.g., a face with a fearful expression vs. a face with a neutral expression). By contrast, seemingly simple features such as shape or color appear not to be discriminable by blindsight patients (assuming, in the case of shape, that the orientation of edges cannot be used to “cheat”).

Thus, within the scotoma of the blindsight patient, there appears to remain a remarkable degree of sensitivity, but it is to a limited set of visual features (like motion and edge orientation). For other features (such as shape, color, or object category), there seems to be little or no sensitivity. The sensitivity of blindsight to certain features but not to others is of great interest in and of itself, as the dividing line appears to map onto known functional divisions in the brain. There is a well-known division of function between the ventral “what” pathway—which proceeds from the occipital lobe along the ventral temporal lobe—and the dorsal “where” pathway—which proceeds from the occipital lobe up into the parietal lobe. The fact that blindsight patients are mostly unable to discriminate form, color, or object category, but are quite good when it comes to movement or position, seems to implicate the “where” pathway. Indeed, the first behavioral evidence for blindsight came in the form of eye movements or reaches toward unseen visual targets.

Neurobiological Explanation of Blindsight

Growing knowledge of brain anatomy and physiology has enabled scientists to offer a plausible explanation for “seeing without seeing” without an appeal to extra-sensory perception or supernatural forces. The most accepted explanation for blindsight (as well as Riddoch’s observations) is the presence of a neural pathway from the retina to the cerebral cortex that bypasses V1. This pathway might carry information without producing any conscious sensation or, in
isolation, produce unusual sensations in response to high-contrast stimuli that move or flicker.

Such a pathway is known to exist: About a tenth of the fibers in the optic nerve project to a primitive midbrain structure called the superior colliculus (the main visual apparatus in frogs and fish) and then, via a part of the thalamus called the pulvinar, to area V5 (also called MT) of the dorsal visual stream. This region of cortex is responsible for processing motion, and damage to this area leads to akinetopsia, or the inability to perceive motion. In one experiment using functional magnetic resonance imaging (fMRI), area V5 was shown to respond to moving stimuli presented in the blind hemifield of patient GY, even when he was unaware that any stimulus had been presented. The pulvinar also projects to the amygdala, an almond-shaped structure in the medial-temporal lobe involved in emotional responses, especially fear. One study reported selective activation of the amygdala in response to fearful faces presented in the blind hemifield of patient GY. Although initially speculative, the existence of a subcortical route to the amygdala is now widely accepted. An even more direct pathway from eye to cortex outside V1 arises from the thin layers of the lateral geniculate nucleus, and, following removal of part of V1 in monkeys, the cortex still responds to visual stimuli presented in the blind part of the visual field via this pathway, as shown recently by Michael Schmid and his colleagues.

Blindsight in Normally Sighted Persons?

Increasing interest in blindsight throughout the 1980s prompted some to search for evidence of “seeing without seeing” in normally sighted persons, but this has proved to be challenging. In the 1990s, two studies, using different methods to manipulate visual awareness, claimed to have found evidence for blindsight in normally sighted subjects. However, one was eventually discounted and the other has been questioned on methodological grounds. More recently, Hakwan Lau and Richard Passingham demonstrated what they refer to as “relative blindsight” in normally sighted observers: Depending on the timing of the backward mask in a visual masking experiment, the proportion of stimuli reported as “seen” can vary even while performance on the task remains the same. The advent of trans-cranial magnetic stimulation (TMS) has allowed researchers to produce a transient scotoma in normal subjects by applying a strong magnetic pulse to the primary visual cortex, synchronized with the onset of each visual stimulus. When this was combined with forced-choice guessing, subjects were indeed better than chance at discriminating certain features of unseen stimuli. It is perhaps not accurate to refer to these as “normally sighted” subjects—the functioning of their visual system at the moment that each stimulus and accompanying TMS pulse were delivered was probably more like that of a patient than that of a normally sighted person. However, the use of TMS does open the door to the possibility of studying blindsight on a much larger scale than has been possible to date. Examples, including the production of experiences described as genuine visual images in the blind region, are given by Alan Cowey in his 2010 review.

Aaron Schurger and Alan Cowey

See also Anosognosia; Subliminal Perception; Unconscious Perception; Visual Masking

Further Readings

Personality disorders are enduring patterns of behavior and ways of interpreting events that are at odds with cultural expectations and lead to significant distress or impairment. Borderline personality disorder entails marked instability of self-image, mood, and interpersonal relationships, along with impulsive, self-destructive behavior. For people with the disorder, perceptions of others tend to swing from idealization to devaluation. Their moods are extremely reactive with episodic severe irritability, anxiety, or euphoria lasting hours to days. They often have inappropriate anger or difficulty controlling their anger. They often suffer from chronic feelings of emptiness and worthlessness and may engage in frantic efforts to avoid abandonment. Potentially self-damaging behaviors such as binge eating, substance abuse, reckless driving, self-mutilation, and suicidality are common. Transient, stress-related paranoid ideation, delusions, or severe dissociative symptoms can occur. Their chronic instability results in marked impairment in functioning. The rest of this entry presents a case study and discusses the disorder’s causes, treatment, and relations to other similar disorders.

Clinical Case Report

Mary is a 19-year-old girl who came to an outpatient clinic for an urgent consultation. She was referred by her primary care physician who found cigarette burns and self-inflicted cuts on her arms. She had recently lost her job in a restaurant because of her irresponsible behavior and irritable mood. She said cutting herself helps her deal with her inner pain. She has had several relationships with men. In each, she falls madly in love, idealizes the person, and in time comes to see him as a terrible person. Mary is preoccupied with a sense of worthlessness and guilt and accepts her sad fate of being chronically unhappy and empty as inevitable.

Etiology

Genetic, psychodynamic, and environmental factors can contribute to the development of borderline personality disorder. Several researchers have proposed the existence of a constitutional incapacity to tolerate stress and regulate emotions. Absence of a warm, supportive parenting figure early in life and traumatic experiences such as abuse can also play a role. As a result of these factors, individuals with borderline personality disorder continue to use the primitive defense mechanisms of splitting and projective identification.

Individuals who use the defense mechanism of splitting are unable to have balanced views of themselves or others. Negative images are so powerful that they threaten to destroy positive ones. Therefore, negative and positive feelings are compartmentalized and shielded from each other. As a result, these individuals alternate between seeing themselves, and others, as all good or all bad, very powerful or weak, loving or mean. Images of the same individual can alternate between these two extremes rapidly. Relationships become very difficult when others are seen as either demons or saints or alternately one and then the other. Moreover, as images of the world descend into darkness, these individuals are filled with despair, engage in drastic attempts to self-soothe (drugs, cutting), or attempt suicide.

In projection, individuals deny that they have undesirable feelings, thoughts, or attitudes and falsely believe that it is others who hold these feelings, thoughts, and motivations. These persons feel that they are not the one who is angry or incompetent or hurtful; rather, it is the other person. In projective identification, individuals not only project their feelings and thoughts but behave toward the other person in such a way that it induces the other person to feel and behave in a way consistent with the projection. In essence, a self-fulfilling prophecy occurs. For example, the individuals deny their own anger and then behave provocatively toward the other person, inducing the other person to become angry and thereby reinforcing the projection. It is very hard for others to absorb the provocative behavior and anger of individuals with borderline personality disorder.

Treatment

Different approaches to therapy have been recommended for individuals with borderline personality disorder. Otto Kernberg suggests a modified psychoanalytic approach to help the individual
Borderline Personality Disorder

resolve pathologic internalized representations of interpersonal relationships.

Others recommend a supportive, reality-oriented approach promoting social adjustment and avoiding regression.

An important aspect of treatment in both of these approaches is that the therapist remains calm and emotionally available, without anxiety or anger. This helps the patient tolerate the hateful and destructive feelings that arise because of transference and to eventually replace them with more constructive and positive reactions. The patient also internalizes a calm, soothing, supportive object.

Dialectic behavior therapy (DBT) is a modification of standard cognitive-behavioral techniques designed specifically for the treatment of borderline personality disorder. DBT focuses on teaching patients four skills: mindfulness (attention to one’s experience), interpersonal effectiveness (predominantly assertiveness), emotional regulation, and distress tolerance without impulsivity.

Brief hospitalization may be needed when the individual becomes seriously depressed and is at risk for self-harm. Hospitalizations only need to be brief, as these moods are often fairly rapidly ameliorated. Moreover, brief hospitalizations are preferable to long-term hospitalizations, which can lead to significant regression.

Medications can be helpful in the treatment of target symptoms such as anxiety, depression, and labile mood. SSRIs (selective serotonin reuptake inhibitors) can help decrease anxiety and depression. Atypical antipsychotics, including olanzapine, clozapine, quetiapine, and risperidone, can help manage psychotic-like, impulsive, or suicidal symptoms.

Differential Diagnosis

The differential diagnosis for borderline personality disorder includes attention deficit hyperactivity disorder (ADHD), post-traumatic stress disorder (PTSD), bipolar disorder, depression, post concussive syndrome, and a difficult or prolonged adolescence.

Individuals with bipolar disorder shift from euphoria or irritability to depression and despair, as do individuals with borderline personality disorder. The speed and reason for the switches are very different, however. In individuals with borderline personality disorder, alterations in mood are much more rapid and are related to events in the environment.

Post concussive syndrome can lead to problems with anger and impulsivity, and individuals with ADHD are impulsive. As a result, on first glance, individuals suffering from these issues may look as if they are suffering from borderline personality disorder. They do not, however, suffer from a sense of emptiness, extreme anger, marked changes in perspective from devaluation to idealization of others, intentionally self-destructive behaviors, and frantic efforts to avoid abandonment, unless they also suffer from personality pathology.

Individuals suffering from PTSD share many symptoms with individuals suffering from ADHD, bipolar disorder, depression, post concussive syndrome, and borderline personality disorder. Marked dysphoria, impulsivity, and transient dissociative symptoms are found in both PTSD and borderline personality disorder. The difference is that PTSD follows a traumatic event about which the individual has intrusive recollections. Borderline personality disorder may be the result of early childhood abuse.

Roy Lubit

See also Anxiety Disorders; Narcissistic Personality Disorder; Self, Development of

Further Readings


First described by Jean Marie Joseph Capgras and Jean Reboul-Lachaux in 1923, the Capgras delusion is one of several rare psychiatric disorders whose symptoms share a common theme—delusional misidentification. It is a universal condition that affects both sexes, and strikes at all ages from adolescence upward. A typical belief held by a Capgras patient is that at least one close family member or friend (some significant other) has been replaced by a physical duplicate—an impostor. This entry presents a brief overview of some of the key features and explanations of the Capgras delusion, including early psychoanalytic and more recent neurological and phenomenological accounts.

Within the Capgras patient, there occurs a conflict of recognition: a seeming paradox in which the significant other is and yet is not recognized by the patient. Physically, including all mannerisms and voice patterns, the impostor is said to be just like the real person yet is not recognized as being that person. The specificity of the delusion is a central feature of the condition and a key aspect to be explained. Early psychoanalytical accounts posited the patient’s putative ambivalence toward the significant other as an underlying cause of the delusion. To cope with this ambivalence, the patient splits the significant other into real and impostor personas. The “real” person symbolizes the ideal—the way the patient should feel toward the other. The “impostor,” on the other hand, provides the patient with a target he or she can safely direct feelings of hostility toward. The delusion is therefore a way of resolving the love-hate conflict underlying the patient’s relationship with the significant other.

In contrast, early theorists proffering a neurological basis for the delusional state noted the prevalence of brain disease or head trauma among the patient group. However, it was initially difficult to understand how organic disease or trauma could account for the high degree of selectivity evident in these delusions. To illustrate, consider Broca’s aphasia, which affects the production of language and is caused by damage to a specific region of the left hemisphere (Broca’s area). How strange it would be if difficulty in language production were directed solely at a select few, with speech occurring normally at all other times and toward all other people.

In 1990, Hayden Ellis and Andrew Young published their mirror-image model that, they claimed, was able to account for the selectivity of the Capgras delusion while positing a neurological rather than psychoanalytic basis for the condition. Ellis and Young argued that the neurological dysfunction underlying prosopagnosia (the inability to recognize faces) is mirror reversed in the Capgras delusion. Research has demonstrated that patients suffering from prosopagnosia, despite failing to consciously recognize familiar faces, nevertheless exhibit increased autonomic arousal when the face is presented. The increase in autonomic arousal is interpreted as unconscious recognition. In contrast, the neurological pathway responsible for conscious recognition, which is damaged in the prosopagnosia patient, is intact in the Capgras patient, and the pathway responsible for unconscious recognition is
damaged. As such, when presented with a familiar face, despite consciously recognizing it, the patient exhibits no increase in autonomic arousal.

The absence of autonomic arousal when in the presence of the significant other is said to pervade consciousness as some form of anomalous, often deeply disturbing, experience. Consequently, when in the presence of the significant other, something feels wrong—there is physical recognition in the absence of emotional connectedness. The stronger the emotional connection between patient and “other,” the more salient its absence is going to feel. Brendan Maher conjectured that the delusional belief is indicative of the patient’s reasoned attempt to make sense of this strange state of affairs.

In contrast, proponents of a two-deficit account argue that the delusional belief (second deficit) stems from the patient’s misinterpretation of the anomalous experience (first deficit). This misinterpretation is the result of some form of cognitive dysfunction—what Max Coltheart refers to as a disruption in the patient’s belief evaluation system. This disruption is thought to be responsible, then, not only for belief formation but, importantly, for its maintenance in the face of overwhelming contradictory evidence.

The two-deficit account has traditionally emphasized the second-stage cognitive dysfunction when accounting for the delusional nature of the belief. While the exact nature of this disruption may vary from theorist to theorist, it is nevertheless commonplace for the patient’s underlying experience to be described simply as anomalous, odd, or bizarre. Yet it is precisely this bizarre experience that the patient tries to explain by mistakenly adopting a delusional impostor belief.

For precisely this reason, Garry Young has recently challenged what he claims is the unidirectional nature of two-deficit explanations. Instead of presenting an account that moves unidirectionally from neurological deficit to anomalous experience to delusional belief, he proposes an amended interactionist model. According to this model, the formation of the delusional belief actually alters the nature of the patient’s experience—moving from an initial sense of estrangement to a full-blown impostor experience. This interaction helps explain why the delusional belief is maintained: The (delusional) belief offers the best explanation for what is experienced because it, in part, helps shape what is experienced. From the patient’s perspective, the congruence between belief and experience provides strong evidence for the validity of the former and the authenticity of the latter.

The neurological dysfunction posited by Ellis and Young’s mirror-image model provides an organic basis for the selectivity characteristic of the Capgras delusion. More contemporary explanations tend to be grounded on this model and differ only in terms of the type of second-stage cognitive deficit they propose in accounting for the formation and maintenance of the delusional belief.

Garry Young

See also Delusions; Face Perception; Fregoli Delusion; Unconscious Perception

Further Readings

CASE-BASED REASONING, COMPUTATIONAL PERSPECTIVES

People who encounter new problems are often reminded of similar prior problem-solving episodes, and those “remindings” can provide starting points for generating new solutions or warnings of potential problems. Remindings can also support interpretative tasks by providing comparison points to compare and contrast to new situations. Computational models of case-based reasoning (CBR) model both of these processes. This entry introduces the CBR process, its tenets, and its ramifications.

Historical Background
The roots of computational models of CBR date to research currents around the early 1980s, including Roger Schank’s studies of reminding and memory, Edwina Rissland’s studies of legal reasoning, and Bruce Porter’s studies of medical diagnosis. In
treating specific stored experiences as the primary knowledge source and the process of adapting prior cases to new needs as the primary reasoning process (rather than the chaining together of general rules), CBR contrasted with rule-based artificial intelligence (AI) approaches. The study of CBR was appealing, both from a cognitive science perspective, for modeling human reasoning, and as a way to further the development of robust and efficient AI systems for reasoning and learning. Reasoning from prior cases may be possible even in poorly understood domains for which it is impossible to account for why the prior case’s solution was successful, and CBR may speed up problem solving by reusing prior effort when generating a solution from scratch would be expensive while maintaining the flexibility to adjust for new circumstances.

Types of Case-Based Reasoning

Case-based problem solving characterizes how specific prior solutions are applied to solve new problems. For example, a doctor deciding how to treat a patient with unusual symptoms might be reminded of the successful treatment of a prior patient, a travel planner might develop a new itinerary by adapting an old one, or a labor mediator might start from a prior labor agreement when seeking a new one. Because CBR may use varied criteria to determine which cases to retrieve and to adapt prior solutions, CBR is not limited to straightforward reuse and may lead to creative solutions. For example, an architect may combine aspects of prior designs—or may use information about those designs to select contrasting features—to creatively design a novel building. Case-based models are also applied to interpretation tasks, to classify or analyze new instances by comparing and contrasting them to prior examples. For example, a travel agent assessing whether a client would like a hotel might do so by comparing it to the most similar hotels the client liked or disliked in the past.

The Case-Based Reasoning Cycle

Agnar Aamodt and Enric Plaza characterize the case-based problem solving process as one of retrieval, reuse of an old solution, revision of that solution, and retention of the result. This process forms a closed loop—from retrieving an old case to storing a new case for future retrieval—and is known as the CBR cycle. Each step in the cycle involves research issues, such as how cases are organized in the memory of cases (called the case base), what retrieval algorithms are used, how solutions are adapted to fit new problems, and how the cases are stored.

The long-term success of systems applying the CBR cycle depends on two types of regularity, one concerning the relationship between problems and solutions in a domain and the other concerning the problems a system will encounter. Problem-solution regularity, the property that similar problems generally have similar solutions, is required to ensure that similar prior cases retrieved by the reasoner will provide useful starting points for new reasoning. Problem-distribution regularity, the property that a reasoner will tend to encounter similar problems over time, is required to ensure that the cases learned by the reasoner will be useful in the future.

Relationship to Other Approaches

CBR contrasts with “eager learning” approaches such as decision trees and neural networks, which form generalizations from training data and then reason from those generalizations rather than from the data itself. Instead, CBR is a “lazy learning” method, which retains raw examples and does further processing only when needed. Case adaptation may be seen as a type of “just-in-time” generalization, done in the context of a new problem and only to the extent needed to handle the new situation.

CBR may be taken as a stance toward cognitive science. In the CBR stance, both commonsense and expert reasoning are seen as largely based on experience, and expertise arises from acquiring cases and the knowledge required to apply those cases effectively. A number of psychological studies provide support for the human use of CBR, and the CBR stance suggests viewing the retrieval, adaptation, and storage of prior experiences as the core of cognition. Computational models of CBR developed in support of this stance provide precise specifications of testable hypotheses of how human CBR functions and the knowledge it requires, addressing questions about memory organization, retrieval, analogical mapping between old and new cases, adaptation of previous solutions, storage, and forgetting. Computational models of CBR have proven useful in artificial intelligence as a basis for studying intelligent systems more generally.
Case-Based Reasoning Applications

Human problem solvers routinely solve problems in domains that are hard to codify precisely. They make reasonable conjectures and learn useful things from both successes and failures, from few examples, and despite limited and uncertain knowledge. The desire to capture these abilities in artificial intelligence systems motivates the development of many CBR applications for tasks such as diagnosis and planning. Especially prevalent are conversational CBR systems, in which cases are retrieved in an interactive dialogue with the user for tasks such as diagnosis. Case-based models of human cognition and CBR applications combine in CBR applications for the learning sciences, which apply CBR to the development of effective teaching strategies and build educational computer systems using CBR.

David Leake

See also Analogical Mapping and Reasoning

Further Readings


Categorization, Neural Basis

Within the cognitive sciences, categorization is defined as the act of responding differently to objects or events in separate classes or categories. It is a vitally important skill that allows us to approach friends and escape foes, to find food and avoid toxins. Not surprisingly, the scientific study of categorization has a long history. For most of this time, the focus was on the cognitive processes that mediate categorization. Within the past decade, however, the new tools of cognitive neuroscience have been used to investigate the neurobiology of categorization processes. As discussed in this entry, the consensus from this work is that all of the major memory systems probably contribute to category learning and that the neural circuits that mediate initial category learning are different from the circuits that enable us to respond automatically to highly learned categories.

Multiple Category Learning Systems

One recent discovery, which is due in part to this new emphasis on neuroscience, is that humans have multiple category-learning systems. An obvious hypothesis, which quickly followed this discovery, is that all major memory systems are capable of some form of category learning. Memory researchers have identified a number of unique human memory systems, which are commonly divided into two classes. Declarative memories are those that are accessible to conscious awareness, such as short-term or working memory and the memory of past episodes (episodic memory). Nondeclarative memories are those for which we have little conscious awareness. Included in this set are procedural memory (e.g., muscle memories, such as the exact actions performed when knotting a tie) and perceptual memories that result from repeated exposure to the same stimulus (i.e., the perceptual representation memory system). Different memory systems have different properties; therefore, each should be ideally suited to learning about unique types of category structures. As a result, changing the nature of the categories might change which brain areas mediate the learning.

Many brain areas have been implicated in category learning, but perhaps the two most important are the prefrontal cortex (PFC) and the striatum. The PFC, shown in Figure 1, is the anterior portion of cortex that lies behind the forehead, and the striatum (which includes the caudate nucleus and the putamen), shown in Figure 2, is a major input structure within a large collection of subcortical nuclei called the basal ganglia.

The Role of the Prefrontal Cortex

A huge literature implicates the PFC in working memory and other executive processes, so the memory systems hypothesis predicts that the PFC should be especially important in tasks where the categories can be learned using executive reasoning. One example is the rule-based category-learning task, in which the categories can be learned via an explicit
Figure 1  A human brain with the prefrontal cortex in a darker shade (the front of the brain is on the right)

Figure 2  The human striatum
There is now overwhelming evidence that the PFC is critical in rule-based tasks. For example, impaired performance on a well-known rule-based task (the Wisconsin Card Sorting Test) is among the most classic of all signs of PFC damage. In addition, many animal lesion studies have confirmed the important role played by the PFC in rule learning and use, and virtually all neuroimaging studies of rule-based category learning have reported task-related activation in the PFC.

**The Role of the Striatum**

The striatum seems especially important to categorization since it has been identified in virtually every cognitive or behavioral neuroscience study of category learning. For example, the striatum is damaged in several neuropsychological disorders, including Parkinson’s disease and Huntington’s disease, and many studies have shown that both of these patient groups are impaired in category learning. In addition, the striatum is frequently identified in neuroimaging studies of category learning, and single-cell recording studies in monkeys have identified neurons in the striatum that learn to respond to the category membership of a stimulus. Finally, a long series of lesion studies in rats and monkeys supports the hypothesis that the striatum is both necessary and sufficient for visual discrimination learning—that is, for learning to emit one response to one stimulus and a different response to some other stimulus. Technically, this is a special case of categorization in which each category contains only one exemplar.

The striatum seems to participate in a wider variety of categorization tasks than even the PFC. One possibility is that whereas the memory-related functions of the PFC seem to be limited to declarative memory systems, such as working memory, the striatum contributes both to declarative and non-declarative memory. For example, a wide variety of evidence implicates the striatum in working memory—presumably because parts of the striatum are reciprocally connected to the PFC. But the striatum is also critical for procedural memory. Historically, procedural memory has been associated with motor skills, such as one’s muscle memory for how to play tennis or golf, but recent work suggests that procedural memory is heavily used in certain kinds of category learning that require integrating information from multiple perceptual attributes in a way that is not easy to describe verbally. One example of such an information-integration task might be learning to discriminate between wolves and German Shepherds. Adults are very good at this task, but they are poor at describing the rule or strategy that allows them to achieve this success.

People learn information-integration and rule-based categories in different ways. For example, information-integration category learning requires immediate feedback after each response, whereas rule-based learning does not. In contrast, feedback processing requires attention and effort with rule-based categories, but not with information-integration categories. There is also good evidence that the striatum plays an important role in information-integration learning. For example, several studies have reported that Huntington’s and Parkinson’s disease patients are both impaired in difficult information-integration tasks. In addition, all known neuroimaging studies of information-integration category learning have reported significant task-related activation in the striatum.

In summary, one reason the striatum may be so frequently implicated in category learning is that it plays a key role in both working memory and procedural memory. A second related reason, however, is that the striatum may be the only brain region capable of reinforcement learning—that is, learning in which synapses active on trials when correct feedback is given are strengthened and synapses active on trials when error feedback is given are weakened. The training signal that makes reinforcement learning possible in the striatum is thought to be provided by the neurotransmitter dopamine, which has two key properties that are critical for successful reinforcement learning. First, much evidence suggests that dopamine levels in the striatum are elevated above baseline following feedback that a response was correct and depressed below baseline following feedback about an error. Second, active synapses are strengthened when dopamine levels are high and weakened when dopamine levels are low. Reinforcement learning is not used to remember a conversation from earlier in the day, because we can remember that conversation regardless of whether it was rewarding. However, reinforcement learning is probably required for a radiologist to acquire
expertise in deciding whether a mammogram shows evidence of a tumor.

The Perceptual Representation Memory System

One memory system thought not to depend on the striatum is the perceptual representation memory system (PRS), which is a nondeclarative system that facilitates perceptual processing of a stimulus as a consequence of having seen that stimulus before. Behavioral effects of the PRS can be observed after only a single stimulus repetition, and PRS effects can be induced when two stimuli are different but perceptually similar. The PRS does not provide a detailed memory trace of an event. Rather, it is thought only to provide a nonspecific feeling of familiarity. Thus, the PRS could assist in categorization only in tasks where exemplars from the contrasting categories are associated with different levels of familiarity.

In (A, not A) category-learning tasks, participants decide whether a stimulus is or is not in the target category A. In this task, the Category A members have a coherent structure since they are created from a single prototype, but typically, all “not A” category members are visually distinct. One might expect more PRS activation on A trials than on not A trials, and in fact, there have been several proposals that the PRS mediates much of the learning that occurs in the (A, not A) task. Neuropsychological support for this hypothesis comes from studies showing that a variety of patient groups with known deficits in rule-based and information-integration category learning are apparently normal in (A, not A) tasks (e.g., Parkinson’s disease patients). The neural basis of the PRS is still unclear, although the evidence is good that PRS effects can be seen in visual cortex. As might therefore be expected, neuroimaging studies of (A, not A) tasks have all reported learning-related changes in visual areas of occipital cortex.

Other Brain Areas

Another brain area that plays a critical role in memory function is the hippocampus. Damage to the hippocampus causes anterograde amnesia, which is an impaired ability to remember recent episodes. For this reason, one might expect the hippocampus to be critical for category learning. Although a few studies have reported slight category-learning deficits in patients with anterograde amnesia, many more have somewhat surprisingly reported near normal category learning in such patients, even though the patients frequently have no memory of any specific category exemplars. Thus, at this point, the exact role of the hippocampus in category learning is unclear.

The sensory association regions of cortex are also important for categorization. For example, infero-temporal cortex (i.e., lying behind the temples), which is critical for high-level processing of visual objects, has drawn interest because of reports that damage in this brain region can cause a category-specific deficit (i.e., an agnosia) in the ability to recognize exemplars from some specific category (e.g., tools or fruits). The most widely known of such deficits occurs with human faces (i.e., prosopagnosia). Despite such results, the evidence is good that category learning does not occur in inferotemporal cortex. For example, categorization training does not make inferotemporal cortex neurons more likely to respond to other stimuli in the same category or less likely to respond to stimuli belonging to a contrasting category (e.g., unlike some cells in the striatum). Similarly, the firing properties of cells in inferotemporal cortex do not change when the category membership of specific exemplars is changed.

The best evidence suggests that although inferotemporal cortex does not mediate the learning of new categories, it is crucial to the categorization process because it encodes a high-level representation of the visual stimulus.

Automatic Categorization

So far, this review has focused on new category learning. There are many reasons to believe, however, that the neural circuits that mediate category learning are different from the circuits that mediate automatic responding to highly learned categories. For example, many neuropsychological groups that are impaired in category learning (e.g., Parkinson’s disease patients) do not lose old, familiar categories (e.g., furniture).

A number of recent studies suggest that whereas the PFC and the striatum are critical for initial learning, the role of both of these areas greatly diminishes after extended periods of practice. For example, there is evidence that after much training, the relevant areas of premotor and motor cortex that mediate the categorization response begin to respond before the PFC or the striatum. Such results are consistent with recent proposals that the development of categorization automaticity is characterized
by a transfer of control from the neural circuits that mediate the learning to cortical-cortical projections between sensory areas that mediate the perception of the stimulus and premotor/motor areas that mediate the motor response. According to this view, a primary role of the striatum in information-integration tasks is to train these cortical-cortical projections.

F. Gregory Ashby

See also Categorization, Psychological Perspectives; Category Learning, Computational Perspectives; Classical Theory of Concepts; Memory, Neural Basis

Further Readings

**Categorization, Psychological Perspectives**

Categories are classes of items that are treated as equivalent with respect to some purpose. We have categories for such diverse items as cars, cheeses, dogs, football games, mothers, and religions. Categories do not simply group items; they provide a crucial way by which we can organize and make use of our experience. We do not need experience with a specific car to drive it or a specific dog to decide whether we can pet it or not. Categorization refers to the set of processes that are involved in putting items into classes and accessing knowledge about the class. It is a central cognitive activity in that it provides a means for going from an observed (or thought about) item to knowledge that is relevant for dealing with that item. It is a remarkable, often effortless, ability that helps our survival. This entry provides an overview of categorization with an emphasis on its various functions and its importance across a wide variety of cognitive activities.

**Functions of Categories**

Categorization allows one both to determine the type of thing one is dealing with and to access knowledge that might be relevant to how to deal with it. Classification, the process of assigning category membership, is clearly a very important function of categories. We need to know if the item is a hammer or a rattlesnake. However, classification is rarely the goal. Rather, we use the classification to appropriately deal with the item. If it is a hammer, we can use it for some purpose, such as putting in a nail or as a paperweight, whereas if it is a rattlesnake, our knowledge suggests moving away.

These other functions beyond classification include just about all cognitive processes, but it is useful to highlight four in particular: prediction/inference, understanding/explanation, reasoning, and communication. First, when we know what category an item is in, we can make some inference about a property or prediction about the future that may affect our plans and actions. Is this dog likely to harm us? Will it make noise? We can use our knowledge about dogs to make more accurate inferences and predictions. Second, we can use knowledge about categories to understand or explain events. Our knowledge about dogs helps us understand why it might be barking and explain why it jumped all over the child who just ate dinner. Third, we can reason about categories based on their knowledge and relation to other categories. If we know a dog is a mammal, we can reason that it is warm-blooded even if we never thought about that before. Fourth, to the extent that people have similar categories, we can communicate with others and learn indirectly about categories. Our knowledge of rattlesnakes, for many of us, is all learned through communication, and we can pass this information to others without the need for them to experience a live rattlesnake.

**Categorical Knowledge for Classification**

A central question is how knowledge about categories is organized to support these various functions.
Much of the work has focused on structure supporting classification, and we examine the major classification views, similarity-based and theory-based, before broadening the discussion to consider other aspects of categorization.

**Similarity-Based Views of Classification**

A simple intuitive idea is that we determine the category of a new item by choosing the category it is most similar to. This new animal is a dog because it is similar to the category of dogs. That is, each item consists of a set of features (e.g., furry, four-legged), and we categorize by finding the category with the most similar set of features. We may distinguish different ways in which similarity is used.

The *classical view* of categories, dating back at least to Aristotle, is much like a definition or rule: All members of a category are similar in that they each have a particular set of features. In addition, any item that has these features is a member of the category. A square is any closed two-dimensional figure with four equal sides and equal angles, and any such figure is a square. This view allows variations in other features, such as size of the square, but ensures a commonality among all category members.

The *prototype view* provides a very different idea of how to use similarity to structure categories, an idea which also points out a problem with the classical view. This proposal is that while each category can be viewed as consisting of similar members, the features do not all have to be true of every member. Rather, category members are organized around a prototype, or best member, that contains all the category features. Other members vary in how many of these features they contain. To classify a new item, one compares it with the prototypes of all the different categories, giving more weight to the important features, and chooses the most similar prototype. Members with more features are viewed as more typical than other members; for example, a robin is a typical bird. Even atypical members, such as a penguin, are likely to match its category prototype more than other categories’ prototypes. You can think of this as related to a family resemblance in an extended family—there may be some family characteristics that one sees when viewing a picture of the whole family, but family members differ in how many of these they each might have. This view avoids the strict definitional requirements of the classical view and may fit better how items from natural categories are classified.

The *exemplar view* also uses similarity in a probabilistic way, but rather than having a summary representation (prototype), this view assumes that people classify new instances in terms of their similarities to specific category members. That is, a new item reminds us of various earlier items and we assume that it is in the same category as those items. Thus, a new large friendly animal may remind one of the neighbors’ large dog and leads to its classification as a dog. A less friendly smaller animal might remind one of a different dog but still lead to the same final classification. Although one might retrieve some items from other categories, for most new items, the majority of retrieved items are likely to be from the same category. Although it may seem strange not to have a single category representation, this view allows much greater flexibility for classification of atypical items.

All of these views have some intuitive appeal and seem to capture some aspects of classification well. Some current *mixed-model* views combine the classical view with one of the probabilistic views to allow rules when useful and more flexible classifications otherwise.

The *theory view* proposes that categories are organized by our theories—explanations regarding relations between categories and/or features of a category. These theories are knowledge based, meaning they are dependent on what we already know. For instance, when learning about a new type of motor vehicle it may be helpful to relate them to other categories such as cars and motorcycles.

A central question in categorization is, What makes categories coherent (i.e., make sense)? The
theory-based view addresses this question. Categories are coherent if they are consistent with one’s knowledge and theories about the world. A category can be considered more coherent if many of its features are causally related to each other (e.g., wings enable flight). Features that are causally related are most likely correlated with each other and will appear together more often than not. Hence, featural similarity is a common, though not necessary, by-product of category coherence. For example, many members of the dog category do look somewhat similar. The important point, though, is that category members also seem similar because they are grouped by some underlying explanatory principle, such as a biological theory (e.g., DNA) for dogs. This idea explains why people are easily able to accommodate category members that do not look very similar to other members, as long as they fit in with the underlying explanation for that category. For example, we understand that beanbag chairs are chairs (though they do not look like other chairs) and that whales are mammals even though they look more like fish.

Even though the theory-based view goes beyond similarity to address why categories seem coherent, the view is complex and not fully developed. It is possible that an underlying explanation does not always drive categorization. In these cases, similarity could be playing a more central role in category organization, and something like the prototype view may be sufficient.

Goal-Based View of Categories

As stated at the beginning, we categorize items with respect to some purpose. We discuss three cases of the importance of this purpose for understanding categorization and then address what this might say about categories in cognition.

Ad hoc Categories

Try to think of the following category: items to take out of your house during a fire. You will notice that it is fairly easy to decide on most items and that the items differ tremendously (e.g., photographs, pets). Ad hoc categories are ones spontaneously constructed in the service of a goal. The basis for this category is not definitional, similarity to a prototype, or similarity to other category members, but rather similarity to some ideal (a highly valuable and portable object). Thus, items can also vary in how typical members they are, in terms of how well they approach the ideal. For example, for the take-in-a-fire category, typical items include baby and money, whereas less typical are favorite clothes and an antique lamp. With repeated use, we may include such categories in our knowledge (e.g., foods to eat on a diet).

Basic Levels and Cross-Classification

Although we often think about items as being in a single category (Fido is a dog; that object is a car), all items belong to many categories. There are two major types of multiple classifications: different levels and cross-classifications at the same level.

All items are part of some hierarchy and can be classified at different levels, such as an item being a Jonathan apple, an apple, a fruit, and a physical object. Even so, there is often a particular preferred level at which items tend to be classified—as an apple, not a fruit or Jonathan apple. This psychologically privileged level is called the basic level. Not only do people usually name items at this level; they are also faster to verify an item at this level (shown a picture, the time to agree it is an apple is faster than to say it is a fruit or Jonathan apple), to learn these terms earlier, and to have a variety of other advantages. This basic level seems to represent a good compromise for two different goals of the system: to have the categories be informative and to have them be distinctive. At higher levels, much information is given up (e.g., saying something is a fruit does not say much about how it tastes), whereas the distinctiveness becomes minor at lower levels (e.g., differences between Jonathan and Macintosh apples). Thus, the basic level may be a useful level for many of the goals for which categories are needed.

Even within a level, however, items belong to many categories. An apple is a fruit and a breakfast food. A person might be a professor, a father, a golfer, and a coin collector. Predicting, explaining, and so on may depend critically on what category one thinks is relevant for the task. Thus, these different categories each provide a rich set of knowledge that may be accessed as a function of how the item is classified.

Category Use

We think of categorization as including classification and the use of the category because in many
situations classification is part of a larger goal-related task. Often, we want to do something with the classification that an item is a car, that a person has a disease, or that the math problem is an example of permutations. We might want to drive a car to our errands, decide on a medication to treat the disease, or access the permutations formula to solve the problem. One could view these goal-related activities as separate from classification, except that they influence later classifications. The way we use a category for a particular goal-related task influences the category representation and, thus, later categorizations. If in solving a permutations problem, we see that order is important, that additional knowledge may influence how we classify later problems. If a treatment of a disease has a large effect on a particular symptom, perhaps that symptom becomes more weighted in our later diagnoses.

Categories and Cognition

Categories are classes of items that are treated as equivalent with respect to some purpose. We can walk through the world classifying objects we happen to see without those items being part of any particular purpose, but much of categorization is goal driven. Even mundane actions, such as unlocking a door, require classifications of the key, lock, and door and an understanding of how the action leads to the ability to get through the door. Those goal-related activities influence how we think about the items and even what we notice about them. For example, people know little about the details of coins (such as pennies), despite having seen thousands of them because most of the details are irrelevant for our goal-related activities with the coins. Many of our uses of items are consistent with their default classifications (apple, dog, hammer), but we can also classify them differently when the purpose of our interaction changes (snack food, comfort, doormat).

Types of Categories

People have categories not only for objects but also for a wide variety of different classes, such as people, ideas, and events. The importance of categorization lies in its usefulness for learning from earlier experiences. By separating different types of things and organizing knowledge around the types, we can apply to a new situation what we learned from earlier situations of that type. We use categories for interacting with people, solving problems, and understanding events. Not only are many of these not objects; many are quite abstract. For example, to speak a sentence, we need to use word classes (e.g., noun, verb), and we understand abstract words such as jealousy or democracy. Some examples provide an idea of categorization’s breadth, while also relating them to goals, similarity-based and theory-based views of category representation, and basic levels.

Social categories are critical for interacting with others. Individuals belong to many social categories, such as student, father, and banker, and these categories convey useful information, such as whether one is likely to be young, tired, or wealthy. We use social categories frequently, and some appear to be activated automatically (gender, age, race). Because people belong to many social categories, which category is most relevant will be influenced by various factors such as context and goals. Social categories help not only to categorize but also to explain social behaviors, to infer unobservable properties, and more generally to shape our perceptions of the social environment.

People often use problem categories to help solve a problem, in both formal education (such as categorizing math problems), as well as in complex real-world situations (such as fixing a car or treating a disease). Categorizing the problem allows one to take appropriate action—to buy a new carburetor or prescribe a strong medicine. Such categorizations are clearly influenced by the goals, and also by the knowledge people bring to bear. Such categories show examples of both similarity-based and theory-based views of category representation across levels of expertise. For example, expert physicists categorize problems using principles and theories involved, such as Newton’s second law, whereas novices rely on superficial properties of the problems, such as pulleys or inclined planes.

Finally, although some categories, such as scenes and events, contain multiple complex objects and relations between them, people quickly identify classroom and beach scenes as well as party events and basketball games. One may categorize these situations at various levels (outdoor scene, beach, white-sand beach; sporting event, basketball game, high school basketball game), but both scenes and events generally have a basic level, just like objects. This level efficiently provides much information relevant to our goals. Being invited to the beach, as
opposed to just the outdoors, suggests what supplies to bring, but knowing that it is a white-sand beach does not add much useful detail.

Brian H. Ross, Eric G. Taylor, and Erin L. Jones

See also Category Learning, Computational Perspectives; Classical Theory of Concepts; Concepts, Philosophical Issues; Conceptual Combination; Human Classification Learning; Similarity

Further Readings

CATEGORY LEARNING, COMPUTATIONAL PERSPECTIVES

Judging a person as a friend or foe, a mushroom as edible or poisonous, or a sound as an l or r are examples of categorization problems. Because people never encounter the same exact stimulus twice, they must develop categorization schemes that capture the useful regularities in their environment. One challenge for psychological research is to determine how humans acquire and represent categories. Formally, category learning can be cast as the search for the function that maps from perceptual experiences to category membership. In this light, various models of human category learning are accounts of how people approximate this function from a limited number of observations. In this entry, human category learning will be considered from this formal perspective.

A function can be seen as a machine that takes inputs and generates outputs. For example, a soda machine (after it receives payment) takes a button press selection as input and outputs the appropriate brand of soda. In algebra class, most students are taught notation for functions, such as \( y = f(x) \) where \( y \) is the output, \( x \) is the input, and \( f \) is the function. For example, \( y = f(x) = 0.5556x - 17.7778 \) is a linear function that takes as input temperatures in Fahrenheit and outputs (i.e., converts to) temperatures in Celsius. Functions can also be nonlinear, such as those that compute compounding interest.

Whereas the temperature conversion function has continuous outputs, category functions have a finite set of discrete outputs. For example, a vertebrate animal can be categorized as a bird, mammal, fish, reptile, or amphibian. Thus, category functions are more like the soda machine than temperature conversion example, though the inputs to the category function can be quite complex, including all that a person can sense.

A Balancing Act Between Flexibility and Bias

Any learning system faces a trade-off that is known in statistics as the bias-variance dilemma. This trade-off involves finding the right balance of inductive bias and flexibility when learning the category function from a limited set of examples (as people do). Inductive bias guides a model’s interpretation of data. To make an analogy, people have an inductive bias to view events co-occurring in time (e.g., smoke and fire) as causally related. A strong bias constrains the form of the category function that a model considers. For example, prototype models are strongly biased to only learn linear mappings (i.e., functions) from stimuli to categories, because prototype models represent categories by a single
average (i.e., abstraction) of category members. For example, a prototype model would represent the category of birds as a single point (i.e., the prototype) that is the average of the features (e.g., size, color) of all birds (e.g., eagles, robins, penguins, sparrows). In practice, prototype models are best for learning categories that have a common family resemblance structure. For example, for the category birds, most birds have characteristics in common—they tend to be small, have wings, can fly, and so on. However, other items violate this structure (e.g., penguins, bats). Thus, the prototype model will have trouble with these items as they go against its bias of categories consisting of one single clump of items. Other models, such as exemplar models, are weakly biased. Rather than averaging items together in memory, the exemplar model stores each item separately, which allows it to learn any possible function. For example, an exemplar model would represent the category of birds as the collection of all birds (i.e., one point for each category member). Exemplar models can learn any category function, whether it be linear or nonlinear. However, even the exemplar model has biases because it will learn some functions more rapidly (i.e., require fewer training examples) than others.

The latter point hints at why biases can be useful. When a model’s bias corresponds to the actual category structure, it will learn the correct category function more rapidly than a less biased model or a model that has an incorrect bias. When a model is too flexible, it will be overly affected by the variance (i.e., noise) of the training examples it observes. In such cases, the model will fail to learn the underlying pattern and be overly affected by the idiosyncratic properties of the examples it has observed. Thus, the category learning function learned will initially be incorrect and will not accurately classify new examples. In general, more flexible (i.e., the less biased) models require more training examples to infer the underlying form of the category function.

Bayesian methods offer a natural way to deal with the bias-variance dilemma by simultaneously considering models of varying complexity (i.e., flexibility). Bayesian methods provide a means to combine prior beliefs with the current observation to determine the probability that an item is a member of a category. Biases for an individual Bayesian model can be explicitly built into the model’s prior (i.e., beliefs held prior to observing any category members), which can be loosely thought of as seeding the model with hypothetical training examples (prior to observing any actual examples). For example, one may have a prior that heads and tails are both equally likely when flipping a coin. Thus, after flipping the coin once and observing tails, one would not conclude that the coin would always come up tails. The prior serves as a bias in favor of certain hypotheses about the category function. As many examples are observed, the importance of the initial prior wanes. This prior knowledge can be quite complex. For example, prior knowledge reflecting biological theories can be incorporated into a Bayesian model designed to learn about animal categories.

When Biases Prevent Learning

The preceding discussion focuses on how biases can be helpful or harmful in promoting (i.e., speeding) the learning of the category function. Of course, very strong biases, such as in the prototype model, can actually make some category functions unlearnable. While researchers typically focus on the rate at which people learn various categories, one important question is whether a model can even learn a category structure. For example, early work in neural networks was criticized and partially abandoned because certain category functions (ones that people could readily learn) were in principle unlearnable by the models. For example, these early models could not learn nonlinear functions, such as exclusive-or (e.g., “a spoon is small and steel or large and wooden”). The learning rules in these models attempt to adjust connection weights to reduce prediction error (i.e., attempt to better approximate the category function). Unfortunately, the models were overly biased toward certain solutions and incapable of learning many category functions, no matter how long the model was trained.

Learnability concerns extend to all modeling approaches. For example, Bayesian models need to be sufficiently flexible (by having a wide range of possible hypotheses about what the category function could be) to be able to eventually learn the underlying category function.

The Curse of Dimensionality

One research challenge is determining what kinds of category functions people readily learn. The size of the input space is a major factor. Problems tend to be easy to learn that involve only a few dimensions.
In the temperature conversion example, the $x$ in $f(x)$ was a single number (i.e., one dimension), as opposed to a lengthy vector of many inputs. As the dimensionality of the input space increases, the number of training examples needed to support learning can increase rapidly. Many learning problems humans face appear to be high dimensional. For example, one could view every receptor in the retina as a dimension for visual learning problems. In practice, these dimensions are not all independent and the brain can take advantage of the structure in the world. Still, one must exploit such biases to learn in large problem spaces. Category learning functions that are smooth and regular are often the most tractable for learning models. Likewise, the effective dimensionality of a category learning problem can be reduced in situations in which attention can be oriented away from certain dimensions. For example, when a mechanic tries to classify what is wrong with a vehicle, the color of the vehicle is usually irrelevant to the decision.

Bradley C. Love

See also Categorization, Neural Basis; Categorization, Psychological Perspectives; Similarity

Further Readings


**Causal Theories of Intentionality**

This entry surveys a range of proposed solutions to the problem of intentionality—that is, the problem of explaining how human thoughts can be about, or be directed toward, objects. The family of solutions described here takes the content of a mental representation—what that concept represents or is about—to be a function of causal relations between mental representations and their typically external objects. This emphasis on causal relations should be understood broadly, however, so as to cover theories couched in terms of lawlike natural relations or the law-governed way in which one natural event carries information about another.

**The Problem of Intentionality**

For good reason, the “aboutness”of human thought seems mysterious, especially to those who embrace the contemporary scientific view of the universe. When a human thinks about his or her next meal, something in the human’s mind represents food or takes food as its object. In contrast, consider the physical relation of being next to. The filing cabinet might be next to the desk, but it is not about the desk; it is not directed at the desk, it does not mean the desk, and it does not take the desk as an object. What in the physical world grounds the aboutness of concepts and mental states?

Causal relations may be the answer. Causes and effects permeate the universe, though, and intentionality does not. Thus, a causal theory of intentionality must identify the particular form or pattern of causal relations that determines the intentional relation.

**Laws and Information**

**Asymmetric Dependence**

In the human case, mental representations most likely take physical form in the neural system, and of course, these neural structures participate in the natural order of causes and effects. This suggests a straightforward causal account of intentional content: A mental representation is about whatever causes its activation (where activation might amount to increased rates of neural firings). For instance, when one sees a cow, this activates the collection of neurons heightened firing of which constitutes the activation of a mental representation—the representation of cows.

We should, however, want to identify a *persisting* representation of cows, one that can be activated in a variety of contexts, on each occasion serving as a vehicle for the subject’s thoughts about cows. This desideratum introduces a complication. We should want standing representations partly because humans...
often think about cows in the absence of cows—that is, even when a cow has not, on the occasion of the subject’s thought, directly caused the activation of any of his or her mental representations. This is especially clear in cases of misrepresentation, in which something other than a cow—say, a horse in the fog—causes a subject to think about cows, because, as we would normally put it, the person has mistaken the horse for a cow. Thus, we require a more discerning causal theory, one that does not simply identify the intentional content of a mental representation with whatever causes its activation. For the simple theory does not seem to allow for misrepresentation; rather, it entails that any cause of the activation of a given mental representation is correctly characterized by the intentional content of that representation.

In response to such concerns, Jerry Fodor develops his asymmetric dependence account of intentional content. The fundamental idea is that relative to the activation of some particular mental representation, certain lawlike processes are derivative on others. There are standard ways in which the activation of a mental representation can be caused, and there are nonstandard ways. Moreover, the former have their status precisely because the other processes depend asymmetrically on them: Were the standard channels not in place, the nonstandard channels would not be either, but not vice versa. The nonderivative law-based relations that cause the activation of a mental representation thus determine its intentional content. A concept represents whatever it is nomically linked to (i.e., linked to by laws of nature) and is such that its nomic link to the concept is the one on which all other such links depend.

Consider a case in which visual input caused by a horse eventuates in the activation of what we normally consider the subject’s mental representation of a cow. According to the asymmetric dependence theory, the mental representation in question would not be activated if it were not the sort of thing activations of which can be caused in a lawlike way by cows. The converse, however, is not true: If the concept in question were to lose entirely its sensitivity to horses, its activation would still be caused by cows. Thus, it represents cows, not horses.

**Informational Semantics**

Alternatively, a mental representation’s intentional content might be the information it carries about some source; that is, the mental representation may simply be about whatever state of the source the mental representation carries information about. Begin with the notion of the amount of information carried. On a specific occasion when a signal is transmitted, the transmitting source is in one of its possible states; so, too, is the device receiving that signal, and this latter state—the state of the receiver—may reveal more or less about the state of the source. If the state of the receiver is consistent with a wide variety of states of the source, then the state of the receiver carries less information about the source than if the state of the receiver had been consistent with only one or two states of the source. As an illustration, consider a case in which an English speaker passes a one-word note to another English speaker. The end of the word is illegible; all that can be made out is *pe*, with a smudge following it. The resulting state of the receiver—the visual apparatus of the person reading the note—is consistent with a substantial range of English words: *pet, pen, percolate, pedestrian*, and many more. Thus, the state of the receiver does not pinpoint the state of mind of the person who wrote the note. In contrast, if the note had shown the letters *perennially* followed by a smudge, the state of the receiver would have carried as much information as possible in this situation; for it rules out all possibilities except that the person writing the note had *perennially* in mind (assuming in this case that the domain of states of the source is limited to thoughts about English words).

The informational approach need not focus only on quantity of information, though. A simple informational theory might hold that the receiver state is specifically about whatever state of the source it homes in on—that is, whatever state (or possible range of states) the source must be in, given the state of the receiver.

Our earlier problem about misrepresentation recurs, however. Whatever the state of the external source, it is thereby among those with which the state of the receiver is consistent! Fred Dretske once proposed to handle this problem by positing a period during which the intentional content of a mental representation is established (and which is then retained by future activations of the mental representation in question). If, during the learning period, a mental representation carries information about only one property or kind, the mental representation is thereafter about that one property or
kind. In contrast, if, during the learning period, the mental representation carries less definite information (its activation is consistent with the presence of more than one possible state of the world), then the mental representation is thereafter about the relevant range of possibilities. Once the learning period ends, the mental representation can be misapplied, thus allowing—as a theory of intentional content should—for misrepresentation.

Causal History

Information and Learning History

Many causal theories take the subject’s history to determine the intentional content of at least some of his or her mental representations. Seeing the shortcomings of the idea of a learning period, Fred Dretske later focused on changes that take place during the learning process itself. Think of an information-bearing structure as a mere detector: When it lights up, it has detected the presence of whatever’s presence is guaranteed by that structure’s lighting up. Such indication can, in some circumstances, provide a reward for the subject. In these cases, behavioral success reinforces the connection between the information-bearing structure in question and the reward-engendering action it caused. As a result, a structure can acquire a function within the subject’s cognitive system—the function of producing the kind of behavior in question. The intentional content of said structure, then, is whatever (a) the structure indicated on the occasion of its acquiring a new function in the cognitive system and (b) is such that the structure’s indicating it explains this modification. Even just one instance of form of behavior can be rewarded, with reinforcement as a result: The mental representation the activation of which caused the subject to exhibit said behavior can now be tightly associated with that form of behavior. Moreover, when this occurs, it can occur because what the activated mental representation carried information about (what it indicated) helps to explain the success of the subject’s behavior on that occasion.

On this view, misrepresentation occurs when a mental representation is applied to something other than that the indication of which explains why the mental representation acquired its role in the cognitive system. Assume, for example, that a neurological structure indicates warmth and, via reinforcement, comes to control, say, certain bodily movements. Developmentally early cases might involve, for instance, the warmth of a parent’s body. Moving toward that warmth rewards the infant or young child by satisfying a desire for, say, human contact; moreover, it does so precisely because that contact comes from the source of warmth. If, at a later time, this mental representation is activated, the child thinks about warmth, regardless of whether, on these further occasions, the representation indicates the presence of something warm or whether moving toward something warm results in a reward.

The Best Test Theory

Robert Rupert’s historically oriented proposal emphasizes comparative probabilistic relations, at least for those representations emerging early in development. The fundamental idea is that a mental representation is about whatever kind or property is the most efficient cause of that mental representation. The efficiency of a cause is measured in the following way. Take a mental representation. For each property or kind of thing that has caused the activation of that mental representation, ask the following: Of all of the mental structures members of that kind (or instances of that property) have activated in a given subject, what proportion were cases of the representation in question? In this fashion, we can ask, relative to a single mental representation (in a single subject), which property or kind is most efficient in its causing of that mental representation relative to its causing of other mental representations. This approach is thus doubly comparative. First, relative to a given mental representation, each kind (or property) has an efficiency rate, which is comparative in the way that relative frequencies are. That is to say, a single kind’s efficiency rate is determined by dividing the number of times it has caused any mental representation at all into the number of times it has caused the activation of the mental representation in question. So its efficiency rate involves facts concerning only the way in which it has caused activation of the mental representation in question relative to its causing of the activation of other mental representations. Second, having in hand an efficiency rate for each kind or property relative to the single mental representation of interest, relations among these efficiency rates determine
the intentional content of the mental representation in question: The mental representation is about the kind or property with the highest efficient rate relative to that mental representation.

Consider a typical subject. Sometimes (on dark nights, for example), cows cause the activation of the representation we would take to be the subject's horse concept, but the efficiency rate of cows relative to that concept is, presumably, very low; of all the times cows have caused the activation of a concept in the subject, the proportion of those that were horse concepts is very low. In contrast, most of the time horses have caused the activation of any mental representation at all, it has been the horse concept, at least for the typical subject. So relative to the horse concept, horses are the winners.

Isomorphism and Teleology

Isomorphism-based views focus on the relation between the internal structure of a mental representation and the internal structure of what it represents: For a mental representation to be about some structure in the world, relations between the elements of the mental representation must mirror the relations between the elements in the thing represented. Moreover, on the explicitly causal version of this view, proposed by Dennis Stampe, a representation's having its particular internal structure must have been caused by the analogous structure in the thing represented. Compare: The elements of a photograph relate to each other in the same way that the elements of the photographed scene relate to each other at the time the photograph was taken.

There is, however, a surfeit of structure in the universe, which leads to a kind of indeterminacy. It may be that at many steps in the causal chain leading to the activation of a mental representation, there appears an appropriate structure—one to which the structure of the representation is isomorphic. Which of the things (external object, structured light, patterns of upstream neural firings) is the object of the mental representation? To solve this problem, isomorphism-based theories typically defer to facts about the purpose or function of various components of the cognitive system—for instance, the visual system's function of tracking objects in the environment.

Many causal theories of intentionality appeal to such teleological considerations in order to resolve indeterminacies or to inform the choice of an intentionality-determining causal relation. Biological theories of intentional contents, so-called teleosemantics, place teleology at center stage. Independent of questions about isomorphism, the general idea is this: The current content of a mental representation is whatever was correlated historically with activations of that kind of mental representation but only in cases in which such correlation explains (e.g., evolutionarily) why mental representations of that type continued to be reproduced (see the work of Ruth Millikan and David Papineau).

Intentional Systems

The theories discussed above assign specific intentional contents to particular mental representations. Perhaps, however, a mental representation has intentional properties only if it appears within a suitable kind of system. For example, it may be that only a system of structures capable of producing intelligent behavior contains elements with intentional content. If there are such further conditions on intentionality, the approaches reviewed in the preceding sections are incomplete; for no physical structure represents simply on account of its satisfying, say, the asymmetric dependence condition. Instead, a structure with intentional content must also appear as part of system with the requisite characteristics.

Robert Douglas Rupert

See also Atomism About Concepts; Classical Theory of Concepts; Intentionality of Bodily Sensation; Intentionality of Emotion; Representational Theory of Mind

Further Readings


Causal theories of memory aim to give a philosophical account of what it is to remember something. According to such theories, to remember something is to be in a mental state or undergo a mental episode that stands in an appropriate kind of causal connection to one or more of one’s previous mental states or episodes. More specifically, causal theories of memory typically have it that the causal connection involves the persistence of a trace: The original experience of an event has left a trace on the subject, which is now operative in the subject’s remembering the event. In psychology, the idea that remembering involves the activation of memory traces is typically taken as uncontroversial, the main focus of research being on what is known as the problem of localization—that is, of trying to identify brain structures that realize such traces. In fact, however, attempting to explain what it is to remember in terms of the idea of a memory trace turns out far from straightforward. This entry summarizes arguments for and against causal theories of memory and distinguishes two ways of understanding the appeal such theories make to the notion of a trace.

The Basic Intuition and the Problem of Deviant Causal Chains
Causal theories of memory typically take as their remit the concept of remembering quite generally, but most of the more detailed analyses focus on remembering particular, personally experienced events, or what psychologists call episodic memory. As applied to episodic memory, one basic intuition that informs causal theories of memory is that there can be cases in which a subject has in fact experienced a certain past event and can now represent it correctly, without it being true that she remembers the event. For instance, a subject might no longer remember her first kiss yet nevertheless represent it correctly because she is reading a diary she kept at the time. The reason why this does not constitute a case of remembering, according to the causal theorist’s diagnosis, is that the subject’s current ability to represent the event does not stand in the right sort of causal connection to her having experienced it.

Obviously, there remains a need for the causal theorist to elucidate what is meant by “the right sort of causal connection” in this context, and what, for instance, disqualifies the kind of causal connection that obtains in our example, where the diary forms the causal link between past and present. This is an instance of the problem of deviant causal chains, which also affects other causal theories in the philosophy of mind (e.g., of perception and action). We can distinguish between two ways of construing causal theories of memory in terms of the type of responses to this problem that they embody.

Memory Traces as Internal
One type of response to the problem of deviant causal chains, in the case of causal theories of memory, involves adding further constraints to the basic idea behind such theories. One such constraint might be that memory necessarily involves traces internal to the body of the subject, which would rule out the diary entry as a suitable trace.

This response turns on an understanding of causal theories of memory, according to which they imply some substantive assumptions about the nature of memory traces, for instance that they must be realized in the subject’s internal physiology (the technology of the day is often invoked to illustrate the idea of a trace in this sense; e.g., wax tablets, phonographic records, or connectionist networks). Yet the more causal theories introduce such substantive assumptions, the less they seem able to offer a plausible analysis of our commonsense concept of remembering, which is typically taken to be their aim. It does not seem true that having the concept of remembering requires having beliefs about physiology, any more than having the concept of talking to someone on a telephone requires having beliefs about the processes that make this possible. Thus, this way of construing causal theories of memory lays them vulnerable to the charge of scientism—that is, of confusing empirical hypotheses about mechanisms underpinning memory with an insight into what it is to remember.

The Very Notion of a Trace
A second type of response to the problem of deviant causal chains, in the case of causal theories of
memory, focuses on the notion of a trace itself and the way it figures in the theory. Going back to our example, there is an intuitive sense in which the diary entry is not a direct trace of the subject’s experience of her first kiss; there is further work for the subject to do, in addition to experiencing the kiss, for the diary entry to be produced. Similarly, it is not the diary entry as such, but the subject’s reading it, that enables her to represent the kiss. So we can draw a distinction on structural grounds between this case and a case of genuine remembering.

Unlike the first type of response to the problem of deviant causal chains discussed above, this response does not introduce substantive assumptions about the nature of memory traces. Rather, it looks at the ontological categories (such as that of an event, an ability, etc.) that we need to invoke in order to get clear about the nature of remembering and understands the notion of a trace as one such category. This way of construing a causal theory of memory can perhaps best be understood by contrasting it with a rival view, according to which remembering should be analyzed in terms of the idea of the retention of an ability to represent the past, where this is explicitly to be contrasted with the idea of the persistence of a trace. Indeed, on that rival view, it is only in cases in which we fail to remember events that we have to rely on traces, as for example when we have to rely on a diary entry to represent our first kiss. One problem with this suggestion is that traces such as diary entries often do play a crucial role in bona fide remembering, by acting as a prompt (or retrieval cue). Consider again the example of a subject reading about her first kiss in her old diary, but not being able to remember the kiss. Contrast this with an alternative scenario in which the memory of her first kiss in fact comes back to her on reading the diary. To spell out what constitutes the difference between the two cases, it seems that we need to introduce the idea that, in the second case, something other than the diary entry grounds the subject’s ability to represent the event. Thus, it is not obvious that there is a viable alternative to causal theories of memory that recruits only the notion of the retention of an ability without presupposing the idea of a memory trace grounding this ability.

Christoph Hoerl

See also Memory and Knowledge; Memory, Neural Basis; Objects of Memory

Further Readings

CHANGE BLINDNESS

Change blindness (CB) refers to the finding that people, in certain circumstances, are surprisingly poor at detecting large changes in visual scenes. This entry will present the methods used to produce CB and discuss how CB findings have informed thinking about the role of attention in scene viewing and about the complexity of visual representations.

In a classic demonstration of CB, observers attempted to detect a change while a photograph of a jet on a runway alternated with a photograph in which the engines had been removed. When a brief (~80 ms) blank frame was inserted between the two images, the change became very difficult to detect, requiring up to 40 alterations between the pictures for successful detection. Once the large change was detected, most people were amazed that they had failed to notice it sooner. This method of producing CB by inserting a blank frame between the two images is called the flicker technique. CB can also be produced when the two images are separated by an eyeblink, an eye movement (saccade), or a camera cut in a video; when an occluding event momentarily blocks the view of the change; when the change co-occurs with the appearance of a number of “mudsplashes”; or when the original scene slowly morphs into the altered image. These varied methods all have one critical aspect in common: When they are implemented, the change fails to produce a clear and isolated motion transient (i.e., a burst of neural activity in response to a change of retinal stimulation). Changes that produce a single clear transient are readily detected, but when people are unable to rely on a transient and must instead rely on their knowledge of the contents of the scene, CB often results.

In the 1990s, these CB findings garnered a great deal of interest. The use of real-world scenes combined with the surprising results revived interest in scene viewing and led to questions about the role
that attention plays in the process. A number of CB findings support the conclusion that change detection in these circumstances requires focal attention. For instance, changes to items that are of central interest in the scene are more easily detected than changes to items of marginal interest. Given that attention should be preferentially allocated to central interest items, this finding is consistent with the view that attention is necessary for change detection. Second, changes become easy to detect if one is cued to attend to the location of the change prior to the change occurring. These findings suggest that attention to an object may be necessary but not sufficient to detect a change. For instance, research finds that people often fail to notice when the main actor in a video is replaced with a different actor during a camera cut. Even in real-life interactions, people often fail to notice when the person they are giving directions to changes to a new individual, provided the direction giver’s view of the swap is blocked by an occluding event. Clearly, in these situations the person who changes is the focus of attention, yet CB results. These observations suggest that one must not only attend to the object in order to detect the change, but must attend to the specific aspects of the object that would allow one to detect the change.

The findings concerning the role of attention in change detection support the view that only attended aspects of a scene are represented with sufficient detail and longevity to allow an observer to readily detect changes. Presumably, these attended aspects are represented within visual working memory and this representation is durable enough to allow a comparison of the prechange item to the postchange item. The conclusion that attention is required for conscious recognition of an object is consistent with conclusions from other methods such as inattentional blindness and the attentional blink.

In addition, the findings that people are often poor at detecting changes seem to indicate that people’s mental representations of visual scenes are far less complete and complex than previously assumed. At the extreme, the findings were used to suggest that people’s subjective experience of a rich and

---

**Figure 1** Example of a flicker change detection paradigm

*Note:* Two images alternate with a blank frame in between them. The participant’s task is to detect the change between the two images (in this case, the striped cushion that appears and disappears on the chair). The display could cycle once or until the participant detects the change; the dependent variable would be percentage correct or time to detection, respectively.
detailed visual representation of the world was an illusion. Instead, it was argued that mental representations of visual scenes were extremely sparse and volatile, changing as the focus of attention switched to different items within the scene.

Although CB research was used to promote this view, other researchers suggested that the CB findings may underestimate the amount of visual information that is represented. This position is supported by evidence demonstrating that people are implicitly aware of changes, suggesting that changes are represented at some level, even if that level is not able to support conscious report. In addition, people began to argue that CB could result from a storage failure or a comparison failure rather than a failure to initially represent information from a scene.

Research following up on these potential shortcomings has demonstrated that people extract more visual information from scenes than CB suggests, storing these visual details in a visual long-term memory of the scene. Thus, CB does not seem to probe the entire contents of visual representations but instead seems to probe the immediately available contents of visual working memory.

As a result, change detection paradigms are currently used to probe the contents of visual working memory and to track the deployment of attention in a scene. In the working memory research, it is assumed that an item is represented within working memory if changes to it are detected. In the attentional tracking research, the time required to find a particular change is used as an indication of when attention was deployed to that aspect of the scene.

Mark W. Becker

See also Attention and Consciousness; Attentional Blink Effect; Inattentive Blindness; Visual Working Memory

Further Readings


CHARACTER AND PERSONALITY, PHILOSOPHICAL PERSPECTIVES

Some kinds of psychological states are dispositions, as contrasted, roughly, with occurrent states, such as a feeling of pain or a perception of a car. Character traits and personality traits are both kinds of psychological disposition. (Quite what dispositions are is a vexed topic on its own, and what psychological dispositions are is equally vexed, as we will see shortly.)

The distinction between character traits and personality traits is not a precise one. Roughly speaking, character traits (such as kindness, benevolence, and open-mindedness) are more evaluative and tend to be more fundamental to an individual’s psychology than personality traits (such as being unsociable and beingumptious). In particular, character traits tend to be more evaluative of a person’s moral worth. The central focus in what follows will be on character traits, not on personality traits or on other kinds of psychological dispositions such as sentiments (being a dog lover) and temperaments (being a melancholic person).

A significant and important challenge to our notion of character traits, and to virtue theory, has come from recent work in social psychology, which can be considered under the general idea of “situationism.” Virtue theory is, roughly, the theory that what matters in morality is more than simply acting well or bringing about good states of affairs but also being a good person, a person with a good character. Situationism will be the main focus of this entry. Before considering situationism in detail, it is important to appreciate that there are in fact two distinct challenges here: (a) a challenge to our ordinary notion of character as embedded in our everyday commonsense psychology and (b) a challenge to virtue theory as found in Plato and Aristotle, and as expressed in much modern philosophical writing in the virtue theoretic tradition. There is a substantial question of what the relation is between virtue theory’s notion of character and the notion of character as it is found in everyday psychology: It is far from clear whether virtue theory is a refinement of
everyday psychology or a correction of it or whether the relation is altogether more subtle and complex. In any event, if the challenge from social psychology to our notion of character in either form goes through, there will be significant implications for moral philosophy.

**Situationism**

A number of findings in social psychology have been brought to bear by psychologists, and adopted by philosophers, to show that our notion of character is not nearly as robust as we take it to be: Subjects’ actual behavior in experimental conditions seems to depend much more on the particularities of the situation than on subjects’ dispositions of character. This leads some situationists to deny outright that there is a viable notion of character. Others more moderately insist that our dispositions to behave are much less stable in their manifestation than we suppose and that they apply across a much narrower range of situation types. So either way, substantial revision in our thinking is required.

Two brief examples will give a flavor of the findings. In one experiment, subjects are located in a phone booth in a shopping mall when they see an experimenter drop her papers; as they see the papers being dropped, some of the subjects “happen to” find a dime in the coin return slot, and others do not. Of those who found the coin, 14 of 16 helped the experimenter pick up her papers; of those who did not (and these subjects have no significant differences in character), only 1 of 24 helped. In another famous experiment, some seminarians stopped to help a seemingly injured colleague and some did not. Whether any particular subject manifested helping behavior depended almost entirely on whether or not he was in a hurry; dispositional facts about him were not a significant variable. Thus, the conclusion is, helping behavior is more dependent on the contingencies of the situation than on character.

While these experiments are no doubt of considerable importance, the fragility of our character traits has been acknowledged, in philosophy and in literature, for a very long time. Aristotle, for example, acknowledged the reality of failures to act as we and others know we should, and the idea in literature of testing a lover’s proclaimed fidelity and finding it wanting (as in social psychology, this is achieved typically through some kind of deception) is a very familiar plot device, from Mozart’s *Così Fan Tutte* to the myth of Cephalus and Procris. And the 1983 film *Trading Places*, directed by John Landis, reveals very nicely how our behavior is, to a considerable extent, sustained by our social role. Nevertheless, it is very important to see these ideas operationalized in a way that is tractable for moral psychology and moral philosophy.

**The Implications of Situationism**

It is generally agreed that our everyday notion of character is open to criticism in a number of respects, and situationism highlights these to considerable effect. First, we do tend to attribute character traits (and personality traits) to people on much too little experience of their behavior across a wide number and range of different situations (e.g., we see someone behave ungenerously just once and then decide that she is an ungenerous person). Second, we often stereotype people, attributing character traits on the grounds of irrelevant other factors (he is from New York, so he is likely to be rude). Third, we tend to take insufficient account of the role in which the person is placed in arriving at our trait attributions (we think that the quiz show host, who knows all the answers, is cleverer than the quiz show contestant, who knows none of the answers).

However, when we turn to the complexities of virtue theory, virtue theorists insist that the right notion of character is much more subtle and flexible than is often presupposed in the social psychology experiments. They insist that character traits are not simply dispositions to behave; to think this would be to conflate character traits with personality traits, many of which are, indeed, understandable in this way (garrulousness is a disposition to talk more than most when in company). What is central to our notion of character according to virtue theory is that a character trait is concerned with the agent’s reasons, and a virtue is a disposition to appreciate all the relevant reasons and to deliberate about those reasons in the right way. For example, generosity is a disposition not just to give to the needy but also to give to the needy for the right reasons. So a generous person would not give to the needy drug addict if it would only worsen his addiction. Only via the agent’s reasons leading to intention and then to action is a character trait to be understood as a disposition to behave. It is in this central respect that
modern proponents of virtue theory argue that situ-

ationism has an overly simple view of what charac-
ter is, and thus, they argue, it misses its target.

**Why Situationism Is Impoverished**

According to Virtue Theory

The central idea in virtue theory of a virtuous char-
acter being a disposition to be properly responsive
to reasons can be brought out in a number of ways.

First, there is the idea that traits are not to be
identified in isolation from each other, for what is
essential to a virtuous person’s reasoning is being
properly appreciative of all the relevant consider-
ations, some of which may be competing so that,
for example, the demands of honesty might prima
facie conflict with the demands of loyalty (or there
might even be competing demands of loyalty—for
example, whether to be loyal to one’s friends or to
one’s political principles). Thus, practical wisdom—
the capacity to appreciate all the relevant consid-
erations—is essential to virtue (as it is not to what
Aristotle called “natural virtue”). And the fact that
someone fails to behave as loyalty might require on
a particular occasion is by no means sufficient to
show that he or she is disloyal or lacking in virtue;
for that person might have been right to have seen
that in this situation and deliberative context what is
most important—what is, all things considered, the
right thing to do—is to be honest.

Second, virtue theorists say that their emphasis
on the agent’s reasons enables them to make the
crucial distinction, which situationism fails to make,
between the person who, for example, acts in a self-

ish way because people are selfish (and to hell with
everyone else) and the person who acts selfishly
because of some kind of temptation or distraction
that leads them to act against their firm resolution to
be generous or to act against their own better judg-
ment on this occasion that it is right to be generous.
The second kind of person, the one who is tempted,
is likely to regret their selfish action, and they might
well try to make amends later.

Third, and related to this second point, virtue
theorists claim that an essential aspect of our char-
acter is the ability to see situations in a certain light
so that, for example, the selfish person, unlike the
unselfish or generous person, will simply not see
someone else’s plight as a situation that calls for
help. But the unselfish people might, if distracted by
irrelevant factors, also fail to see the situation as they
should: For example, they might not see the other’s
distress because they are in a hurry and this again
will be something that they may well later regret.
Situationism, it is claimed, blurs all these vital dis-
tinctions that virtue theory insists on.

**Why Virtue Theory Is in Need of**

**Modification According to Situationism**

The situationist might well accept these points, but
he or she can still insist that virtue theory has a lot to
learn from situationism.

First, the situationist will say that the notion of
character in virtue theory represents an ideal of
character that the experiments in social psychology
have shown to be something that cannot be realized
in human beings. And any ethical theory ought to
be concerned with what human beings are in fact
capable of. (This last point is contentious: The virtue
theorist might reply that we need to aim for ideals
even if we cannot possibly achieve them.)

Second, the situationist will say that virtue
theory operates with far too broad a notion of a
trait, assuming that, for example, honesty will be
manifested across a whole range of different types
of situation, from the workplace to the home, from
marking exams to filling in tax returns. Yet this
cross-situational consistency is belied by a multiplic-
ity of experiments. Moreover, he or she will add, we
not only cannot expect consistency within a particu-
lar virtue; we also cannot expect consistency to be
manifested across virtue; the so-called doctrine of
the unity of the virtues, with practical wisdom at its
core, is quite unrealistic.

Third, the situationist will argue that virtue the-
ory, in its idealization of human character traits, is
deeply problematic in its practical application. First,
with an ideal of virtue in mind, it places insufficient
emphasis on the need to avoid situations that might
tempt those of us who are less than fully virtuous
(i.e., all of us). Second, we not only need to avoid
tempting situations, but we also need to recruit the
help of others to see us through our resolutions. And
third, the ideal of character found in virtue theory
can easily lead to a culture of blame—of holding
ourselves and others irretrievably responsible for our
failures of character, when in many cases these fail-
ures are ones that any of us would have manifested
in the same circumstances.
Further Implications of the Debate

This debate has significant implications not only for philosophical accounts of moral theory but also for moral practice—for moral education, for moral responsibility, and for punishment. All of these implications need to be explored in the light of the empirical findings in social psychology, but, the virtue theorist would add, properly informed by a notion of character that is not simply understood as a disposition to behave in a certain way.

Peter Goldie

See also Attitudes and Behavior; Attribution Theory; Emotion and Moral Judgment; Moral Development; Personality: Individual Versus Situation Debate

Further Readings


Classical Theory of Concepts

The classical theory of concepts is the view that, at least for ordinary concepts, a subject who possesses a concept knows the necessary and sufficient conditions for something falling under the concept—that is, the definition of the concept. This entry first introduces the theory, and then discusses two powerful objections to it.

There are now many alternative theories of concepts on offer, but they all are, in some way or another, reactions to “the classical theory of concepts,” or “the definitional view of concepts.” The classical theory has two different aspects, though they are often distinguished insufficiently. First, it is assumed that all concepts (except, perhaps, some specific basic or primitive concepts; see below) have a classical analysis, or a definition, in terms of simpler concepts—a definition that gives necessary and sufficient conditions for falling under the concept, or, in other words, for belonging to its extension. Second, it is presupposed that if a subject has or possesses a concept—grasps it—then he or she must know the relevant definition or correct analysis.

Consider thus the worn out example, bachelor (italicized expressions are here used to designate concepts that are expressed by the corresponding words and phrases; the latter are mentioned by putting them in quotes; e.g., “bachelor”). Now, according to the traditional view, bachelor is constituted by the more primitive concepts unmarried and man; thus, bachelor can be defined with the latter, or, those concepts provide its analysis. Hence, it is necessary and sufficient for something to fall under the concept bachelor to be unmarried and a man. Moreover, anyone who possesses the concept bachelor must know this. Or to change the example, presumably the concept vixen can be defined in terms of the concepts female and fox and anyone who has the concept vixen must know that this is the case.

In accordance with such suggestive paradigms, it has been thought that virtually all concepts have a definition in an analogous way. For example, tiger might get analyzed along the lines of a large, carnivorous, quadrupedal feline, tawny yellow in color with blackish transverse stripes and white belly; consequently, something would fall under the concept tiger if, and only if, it has the latter properties. Or perhaps lemon can be defined as, say, pale yellow, tart, oval citrus fruit and so forth; it is necessary and sufficient for belonging to the extension of lemon to have these features. And once again, the idea is that a subject would possess the concept tiger or lemon only if he or she knows these definitions.

It is commonplace to use specific Latin terminology in such cases: If it is definitions that are under
discussion, the concept being defined, such as bachelor, is called “the definiendum,” and what is offered as the definition, such as unmarried man, “the definiens.” Analogously, in the case of analysis, what is analyzed is “the analysandum,” and what provides the analysis “the analysans.”

The classical theory of concepts is deeply rooted in the tradition of western philosophy: In Plato’s dialogues, Socrates is described as asking for classical analyses of various concepts; definitions also have an important place in Aristotle’s thought. In early modern philosophy, the classical theory was also dominant; one can find expressions of it in the works of René Descartes, Blaise Pascal, Baruch Spinoza, and especially John Locke, for example. It is also clearly present in the early 20th century in the thinking of Bertrand Russell, G. E. Moore, and the logical positivists and has indeed been a cornerstone of the traditional analytic philosophy, which emphasized conceptual analysis so strongly. In more recent philosophy, Frank Jackson and Christopher Peacocke, for example, have advocated variants of the classical theory.

Apparently, some classical thinkers expected every concept to have a classical definition but on closer scrutiny, such a view is very difficult to defend. As a result, beginning at least with Pascal (in the 17th century), it has been widely granted that some concepts must be primitive and cannot be further defined. The question then arises, though, how are these basic or primitive concepts possessed or grasped? In the rationalistic tradition following Descartes, it has been typical to think that these are grasped with some sort of immediate intuition. Moreover, it has been characteristic for this school to think that these concepts are not acquired at all but innate. Empiricists such as Locke, David Hume, and their followers, in contrast, have always been skeptical about such ideas, and have held that all concepts must be ultimately grounded in sense perception. Accordingly, it has been suggested that primitive concepts are somehow made to correspond to simple sensations or impressions, or “sense data.” Such classical empiricist views of concepts have had considerable troubles in the last half a century or so. The idea of innate concepts, on the other hand, has experienced a new revival, largely because of the works of the distinguished linguist Noam Chomsky and especially of Jerry Fodor, a leading philosopher of cognitive science influenced by Chomsky. Fodor himself does not, however, combine this idea with the classical theory but argues that almost all concepts have no definition but are primitive. In any case, the idea of plentiful inborn concepts remains far from uncontroversial.

Objections to the Classical Theory

Though once the prevailing view of concepts, the classical theory has been under much attack more recently. To begin with, it is an undeniable fact that we have really achieved satisfactory definitions of none too many concepts. In psychology and cognitive science, the classical theory has also been criticized from the direction of competing theories such as the prototype theory. Moreover, it has been argued that a number of vague concepts such as bald and short cause troubles for it. Let us, however, in the rest of this entry, focus on a couple of more general problems of principle.

Quinean Doubts: Definitions as Episodic

There is an argument due to W. V. Quine, and developed further by Hilary Putnam, which suggests that there are many concepts that do not have a definition, in the sense of the classical theory of concepts. The gist of Quine’s argument is that even if a concept is originally introduced into science via an explicit definition, definitions in science are episodic; that is, the status of the resulting equivalence of the new concept and its definiens need not be eternally privileged, a necessary truth, or true by convention. Putnam has developed this view further by introducing his notion of “law-cluster concepts.” These are concepts that are implicated in a number of scientific laws. And if any one of these laws is treated as a necessary condition for the concept, one is, Putnam submits, in trouble.

For example, in the classical Newtonian physics the concept momentum was defined as mass times velocity. Soon after, it became clear that momentum is a conserved quantity. The law of conservation of momentum, in contrast to the definition of momentum per se, was originally treated as an empirical law, but later, it and the above definition functioned on a par. When Einstein later developed his special theory of relativity, it turned out that this theory was in conflict with the assumption that momentum is mass times velocity. It was the latter thought that was revised, however much it was treated as a definition earlier.

Such considerations strongly suggest that, at least for certain kind of scientific concepts, the classical
theory of concepts fails. Note that this line of argumentation does not claim that there are no analytic truths or that *vixen*, for example, could not be correctly analyzed as *female fox*. The point is, rather, that concepts such as *vixen* or *bachelor* are very special (in Putnam’s terminology, “one-criterion” concepts) and not representative and do not offer a good model for a general theory of concepts. Many other concepts do not have any such standing definitions, if the argument is sound.

**Concept Externalism**

A different line of critique derives from “semantic externalism,” developed by Saul Kripke, Hilary Putnam, and others. Though the arguments originally focused on linguistic expressions and their meaning, it didn’t take long before they were also applied to concepts. In this latter development, the work of Tyler Burge has been especially influential. At the core of this approach are various “arguments from ignorance and error.” They aim to demonstrate that the classical view and its kin require, from the average persons, knowledge they plainly do not have and emphasize how fallible and ignorant we all tend to be. Often, they are based on the plausibly general assumption that the underlying reality may go beyond perceptual appearance.

Consider thus, for example, the concept *dolphin* or *dinosaur*. Ignorant persons might suggest that *fish* is a constituent of the concept *dolphin* or that *prehistoric lizard* is a part of *dinosaur*. But in fact, dolphins are not fishes, and dinosaurs are not lizards. The definitions these people would provide would go wrong. Or, let us reflect the above examples, *tiger* and *lemon*, and their proposed definitions. Now whether something really is a tiger or a lemon is in part a matter of “inner nature” (such kinds are standardly called “natural kinds”). Therefore, it is not impossible that a creature had tawny yellow color, stripes, and so on—indeed, was indistinguishable from tigers—and still failed to be a tiger; it might have a radically different inner nature. Similarly, a fruit might well be pale yellow, tart, and oval and nevertheless not count as a lemon, because it has a wrong kind of inner nature. On the other hand, such kinds may have untypical members: A tiger might have only three legs, lack stripes, be very tame, and be a vegetarian. Or some lemons may not be yellow, tart, or oval. Nevertheless, these would count as tigers or lemons, respectively, because they have the right genetic structure and lineage. Hence, the conventional definitions that people would associate with concepts may fail to provide either necessary or sufficient conditions.

The most famous externalist argument is the Twin Earth thought experiment due to Putnam: Imagine that somewhere, there is a planet very much like Earth, “Twin Earth.” Even languages similar to ours are spoken there. There is, however, a difference: The liquid called “water” there is not $H_2O$, but a totally different substance; call it XYZ. It is assumed that it is indistinguishable from water in normal circumstances; it tastes like water and quenches thirst like water, lakes and seas of Twin Earth contain XYZ, and so on. Let us further assume that modern chemistry has not yet developed in either planet. Consequently, nobody would have been able to differentiate between XYZ and $H_2O$, and people in Earth and Twin Earth would associate exactly the same “definition” with the substance they call “water,” along the lines of *liquid, bright, tasteless, quenches thirst, fills lakes and seas*. Nevertheless, their respective concepts, so the argument continues, must be different, for they have different substances in their extension. The best definition people could give would not be sufficient.

The essential idea of externalism thus is that what entities really fall under a concept—say, *tiger* or *water*—may partly depend on external features of the environment unknown by the subject. The soundness of externalism itself is certainly a vividly debated topic, but it is fair to say that even its critics typically give up some essential features of the classical theory of concepts.

Panu Raatikainen

See also Anti-Individualism About Cognition; Atomism About Concepts; Categorization, Psychological Perspectives; Concepts, Philosophical Issues

**Further Readings**


**COGNITIVE DISSONANCE**

Cognitive dissonance theory originally predicted that when an individual has in mind two or more elements of knowledge that are relevant to each other (and important) but inconsistent with one another, he or she will experience a state of psychological discomfort, which is referred to as dissonance. Lying to a friend is just one of many examples of a situation that might create dissonance. Thus, the theory is concerned with how perception and cognition influence and are influenced by motivation and emotion. Over the past 50 years, hundreds of experiments have tested dissonance processes, and research on the theory continues to this day. For the most part, these experiments have explored the ways that the experience of cognitive dissonance causes attitude and behavior changes, although the theory’s reach is much broader. This entry reviews the original theory of cognitive dissonance and the experimental methods used to test it; discusses revisions to the original theory; and surveys recent research on neural circuits involved in dissonance processes, cultural differences in dissonance processes, neural network models, and nonhuman animal responses to dissonance.

The theory’s long and continuing impact probably resulted for several reasons: The theory is stated abstractly, can be applied to a wide array of issues, and deals with the interaction of cognition, motivation, and affect. Moreover, the theory has generated research that suggested ways of inducing lasting attitude, belief, and behavior change. For example, dissonance research has demonstrated that dissonance processes can reduce prejudice, increase water conservation, increase the purchasing of condoms, reduce hunger and thirst, and reduce pain. In addition, dissonance processes can lead to changes in attitudes toward a variety of objects and issues, such as boring tasks, boring reading passages, delicious chocolate, eating grasshoppers, bitter beverages made with vinegar, increasing tuition at one’s university, and mandatory comprehensive final exams.

**The Original Theory of Cognitive Dissonance**

In the mid-1950s, Leon Festinger proposed the theory of cognitive dissonance, which as mentioned above, states that an unpleasant state of psychological discomfort, dissonance, is created when an individual holds two or more elements of knowledge of some importance that are relevant to but inconsistent with one another.

According to the original theory, the degree of dissonance in relation to a particular cognition = \( D / D + C \). In this formula, \( D \) is the sum of cognitions dissonant with a particular cognition and \( C \) is the sum of cognitions consonant with that same cognition. Also, in this formula, each cognition is weighted for importance.

Festinger theorized that the unpleasant state of dissonance motivates individuals to engage in cognitive work to reduce the inconsistency. This can be accomplished by adding consonant cognitions, subtracting dissonant cognitions, increasing the importance of consonant cognitions, or decreasing the importance of dissonant cognitions. Attitude change is one of the most often assessed ways of reducing dissonance. This attitude change is expected to be in the direction of the cognition that is most resistant to change. In experimental tests, researchers often assume that the knowledge about recent behavior is usually most resistant to change, because it is often very difficult to undo that behavior. This concept of resistance to change of cognition is vital to the theory and distinguishes dissonance theory from other theories concerned with similar processes. The concept, resistance to change, specifies which cognitions will likely be altered by the individual experiencing dissonance.

**Experimental Paradigms Used to Test Dissonance Theory**

The theory of cognitive dissonance has been tested using a number of experimental paradigms. One paradigm is referred to as the free-choice or
difficult-decision paradigm. In this paradigm, the individual is asked to make a decision that is difficult. After a decision, all of the cognitions that favor the chosen alternative are consonant with the decision, while all the cognitions that favor the rejected alternative are dissonant. The greater the number and importance of dissonant cognitions and the lesser the number and importance of consonant cognitions, the greater the degree of dissonance experienced by the individual. In a decision situation, dissonance is typically greater the closer the alternatives are in attractiveness (as long as each alternative has several distinguishing characteristics). Dissonance caused by a decision can be reduced by viewing the chosen alternative as more attractive and/or viewing the rejected alternative as less attractive. This effect has been referred to as spreading of alternatives. For example, a person might have to decide where to attend university or simply have to decide what to eat for dinner. If the decision is difficult, the person will value the chosen university more after the decision than prior to the decision; he or she will also devalue the rejected university more after the decision than prior to the decision.

Another paradigm is the induced compliance paradigm. According to the logic underlying it, dissonance should be aroused when a person acts in a way that is contrary to his or her attitudes, as long as he or she is provided little external justification for doing so. One common way of accomplishing this in experiments is to provide individuals with an illusion of choice in engaging in the counterattitudinal behavior. Individuals may reduce this dissonance by changing their attitudes to be more consistent with their actions or by perceiving the counterattitudinal behavior as less important, depending on the resistance to change of relevant cognitions. For example, people might be asked to endorse or write a statement supporting a social policy (e.g., gay marriage) with which they might not agree. If they make the statements with little external justification, they will change their attitudes to become more favorable to the stated position.

Dissonance can also be aroused by exposure to information that is inconsistent with beliefs or attitudes. Typically referred to as the belief disconfirmation paradigm, this situation involves presenting information inconsistent with an individual's firmly held beliefs or attitudes. It is based on Festinger and colleagues' field observations of increased religious fervor among members of an apocalyptic religious group following a predicted cataclysm that failed to transpire. Subsequent studies have supported the theory-derived prediction that when individuals' firmly held beliefs are unequivocally disconfirmed, they often intensify the original beliefs.

Acting hypocritically can also arouse dissonance. When a person says one thing but then behaves opposite to what he or she said or “preached,” that person may experience dissonance. Experiments using this hypocrisy paradigm have revealed that when individuals are reminded of their hypocrisy immediately after advocating a particular position, they will reduce dissonance by acting more in accord with their advocacy. For example, reminding individuals of times they failed to practice safe sex just after they have made persuasive speeches on the benefits of safe sex causes them to purchase more condoms, presumably because they intend to practice more safe sex in the future.

Theoretical Revisions of Dissonance Theory

Some theorists hypothesized that the effects obtained in dissonance experiments were due to nonmotivational, cognitive processes or impression management concerns. However, subsequent research confirmed that dissonance is a motivated process. In particular, research revealed that the dissonant cognitions cause negative emotional states, and these negative emotional states motivate the cognitive and behavioral changes. Following this dissonance reduction, the negative emotional states subside. Beginning in the late 1960s, researchers began to propose motivational explanations for dissonance effects that differed from Festinger’s theory. Four revisions have been proposed.

Elliot Aronson proposed that dissonance is not due merely to an inconsistency between cognitions. Instead, he posited that dissonance occurs when a person acts in a way that violates his or her self-concept—that is, when a person performs a behavior inconsistent with his or her sense of self. Since most persons have a positive self-concept, dissonance is most often experienced when a person behaves negatively, behaving in an incompetent, irrational, or immoral manner. One of the primary predictions derived from this revision is that low and high self-esteem individuals should respond with less and more dissonance reduction (e.g., attitude change),
respectively, because in dissonance experiments high self-esteem individuals are induced to act in ways that are more discrepant from their positive self-views. Experiments testing this prediction have produced mixed results.

Claude Steele’s self-affirmation theory proposed that persons possess a motive to maintain an overall self-image of moral and adaptive adequacy. Accordingly, dissonance-induced attitude change occurs because dissonance threatens this positive self-image. While Festinger’s dissonance theory posited that individuals are motivated to reconcile inconsistent cognitions, Steele proposed that, instead, individuals are merely motivated to affirm the integrity of the self. In support of this idea, Steele presented experiments, where, following a dissonance induction, participants either were, or were not, presented with an opportunity to affirm an important value. When participants were allowed to affirm an important value, dissonance-related attitude change did not occur. Other experiments have suggested that making important but non-self-affirming values salient reduces dissonance by reducing the individual’s perception of the importance of the dissonant act, consistent with Festinger’s theory.

Joel Cooper and Russell Fazio proposed the idea that the discomfort experienced in dissonance experiments was not due to an inconsistency between the individual’s cognitions but rather due to feeling personally responsible for producing an aversive consequence. In support of this idea, experiments revealed that dissonance-related attitude change only occurred in conditions in which an aversive consequence was produced. More recently, experiments have found dissonance-related arousal and attitude change in induced compliance conditions where individuals do not produce aversive consequences.

Several experiments since 1995 have supported the original conception of dissonance theory, but why does dissonance evoke this emotive state? Eddie Harmon-Jones proposed an action-based model of cognitive dissonance in an attempt to answer this question. The action-based model proposes that the perceptions and cognitions likely to arouse dissonance are those that are associated with action tendencies. The action-based model further proposes that dissonance between cognitions evokes an aversive state because it has the potential to interfere with effective and unconflicted action. Dissonance reduction, by bringing cognitions into consonance, serves the function of facilitating the execution of effective and unconflicted action. Experiments have revealed that experimentally increasing the degree of action orientation experienced following difficult decisions increases the degree of dissonance reduction.

Neural Circuits Involved in Dissonance Processes
Physiologically oriented research has revealed that the state of dissonance is associated with activation of a brain region known as the anterior cingulate cortex and with activation of the sympathetic nervous system. Consistent with conceptual models of cognitive control, evidence also suggests that the reduction of dissonance involves activation of the dorsolateral prefrontal cortex, particularly the left dorsolateral prefrontal cortex.

Cultural Influences on Dissonance Processes
Some researchers have suggested that dissonance processes only occur in Western cultures but not in Eastern cultures (i.e., East Asia). Other researchers, however, suggested dissonance-related attitude change can occur in individuals from East Asia when those individuals are concerned about the interpersonal consequences of their actions. That is, individuals from East Asia, who are typically motivated to be interdependent with others and avoidant of interpersonal conflict, experience dissonance when they act inconsistently with those cultural ideals. Along similar lines, research has revealed that individuals experience dissonance even when other individuals with whom they identify behave in ways that would arouse dissonance for the perceiver (e.g., your favorite political candidate behaves immorally). These results serve as a reminder that the importance of the cognitions is one of the factors affecting the magnitude of dissonance in Festinger’s original theory. Cultural values would be expected to relate to the importance of cognitions and, thus, to the amount of dissonance these behaviors would evoke.

Neural Network Models and Nonhuman Animal Data
Several neural network models have been developed to predict how cognitive dissonance will influence an individual’s attitude and behavior. Experiments have also demonstrated that nonhuman animals
Cohort Model of Auditory Word Recognition

The core idea at the heart of the cohort model is that human speech comprehension is achieved by processing incoming speech continuously as it is heard. At all times, the system computes the best interpretation of currently available input combining information in the speech signal with prior semantic and syntactic context. Originally proposed in 1980 by William Marslen-Wilson and Lorraine Tyler, the cohort account has been subject to ongoing refinement in response to new empirical data and neural network simulations. Predictions of the model for neural responses to speech are currently being tested.

Origins of the Cohort Model (1973 to 1985)

During the 1970s, response time data collected by Marslen-Wilson and others demonstrated the speed and accuracy of speech perception and comprehension. Native speakers can shadow (i.e., repeat aloud) heard sentences with minimal delays between perception and production while correcting for mispronunciations of key words. Detection tasks similarly show rapid word identification in sentences with participants responding within 300ms of the start of a target word—substantially before all of the relevant speech has been heard. Early identification is achieved by comparing incoming speech with known lexical items (for example, the word trespass is distinct from all other words once the segment /p/ has been heard) in combination with contextual cues provided by the sentential context.

The cohort model proposed by Marslen-Wilson and Tyler in 1980 thus suggests that word identification begins with initial activation of a set of candidates that match the start of a spoken word (the word-initial cohort that for trespass would include words such as tread, treasure, treble, etc.). Activated candidates are rejected as incompatible speech segments are heard; word recognition occurs when information in the speech signal uniquely matches one single word (the uniqueness point). The importance of the sequential structure of spoken words in predicting word recognition has been confirmed in a range of response time tasks. For example, people can decide that speech does not match a real word (i.e., make lexical decisions about spoken pseudowords) as soon as they hear a segment that deviates from all spoken words. Hence, decision responses occur with a constant delay when measured from the /s/ of the tromsone or the /p/ of trombope; it is at these positions that participants can determine that
these pseudowords are not the familiar word trombone. These findings uniquely support accounts of speech perception in which word recognition can occur when sufficient information has been perceived in the speech signal and provide strong support for the first instantiation of the cohort model.

Computational Instantiations
(1986 to 1999)

The predictions of the cohort model for the timing of word recognition depend on assessing phonetic information that is shared by or distinguishes between different spoken words. These can be estimated from computerized pronunciation dictionaries revealing the proportion of items (like doll embedded in dolphin) that challenge early identification. A further methodological development was the use of cross-modal priming to test the time course of word recognition. Consistent with cohort assumptions, Pienie Zwitserlood showed that speech that matches multiple words (e.g., the start of captain or captive) activates multiple meanings to a degree that is modulated by prior sentence context. Word frequency also affects activation of word candidates, so captain would be more active than capstan because of its higher frequency of occurrence. These findings motivated a revision to the cohort theory in which word activation and identification are graded processes that combine the speech input, lexical information, and contextual cues.

This revised version of the cohort model shares a number of important characteristics with the TRACE computational model of spoken word recognition developed by Jay McClelland and Jeff Elman in 1986. This is a simulation of a network of simple neuron-like processing units (neural network), that simulates the processes by which people recognize spoken words. This model motivated further tests to distinguish between the Cohort and TRACE theories. Two findings that are contra-TRACE are that (a) ambiguous phonemes do not slow down the identification of pseudowords in the same way that they slow down word identification, and (b) mispronunciations that create pseudowords impair recognition (e.g., hearing bandin for bandage has similar effects to cabin for cabbage). However, eye movement effects observed for rhyme competitors (e.g., looking at a picture of a speaker when hearing the word beaker) are more consistent with the predictions of TRACE.

Much of this behavioral data is well simulated by a different neural network simulation of word recognition: the Distributed Cohort Model (DCM) developed by Gareth Gaskell and William Marslen-Wilson in the 1990s. The DCM differs from TRACE both by using distributed representations (i.e., encoding speech information over multiple rather than single artificial neurons) and by simulating phoneme perception and word identification as two parallel processes rather than two sequential stages. This allows the model to simulate the differential impact of ambiguous speech sounds (e.g., a sound that is acoustically intermediate between an /s/ and an /h/) on phoneme and word perception. It also provides the most accurate simulation to date of cross-modal priming results similar to those reported by Zwitserlood. Lexical identification in DCM is achieved by activating output representations in proportion to their conditional probability given the current speech input, a proposal that to some extent anticipates recent Bayesian theories of language comprehension.

Expansion and New Horizons
(2000 to Date)

Following the success of the Distributed Cohort Model in explaining adult recognition, behavioral studies have focused on extensions of this account to new empirical data, populations, and explanatory domains. One focus of recent research is the recognition of words such as darkness and darkly that comprise a stem (dark) and affix (-ness, -ly). These polymorphemic words are decomposed into their constituents during identification and behave differently from morphologically unrelated pairs such as dolphin and doll. A further extension of the cohort model has been to account for behavioral evidence concerning word learning in adults and infancy. In adults, new learned spoken words do not show cohort-like recognition behavior straightaway, but only if a period of sleep follows initial learning.

Infants show adultlike recognition behavior early on in development (under 2 years old) since they also recognize words as soon as the speech signal allows similar sounding words to be ruled out.

Perhaps the most ambitious extension of the cohort account, however, is to incorporate constraints provided by functional brain imaging.
The continuous processing of speech in the cohort account have been validated by electro- and magnetoencephalography (E/MEG) data showing that neural responses to spoken words are timed to the perception of critical segments in speech. Support for separate semantic and phonological pathways as in the DCM has come from brain imaging studies consistent with dual-pathway accounts of the neuroanatomy of spoken language. A challenge for future imaging work is to establish the unique functional contribution and timing of neural responses within these distinct pathways.

Matthew H. Davis

See also Electrophysiological Studies of Mind; Frequency Effects in Word Recognition; Speech Perception; Word Recognition, Auditory; Word Recognition, Visual

Further Readings

**COLLECTIVE ACTION**

Consider people walking in the street. As such, this is not normally regarded as collective action even if there are many agents acting in the same location—some of them are going for lunch, others are shopping, and so forth. By collective action we typically mean action by several agents who are suitably socially connected (e.g., are acting on the shared goal to go for lunch to a certain restaurant or are acting on the belief that there will be a traffic jam because of a sports event). Here, the same type of action is being performed with the knowledge that the others are also performing it or a related type of action. This entry discusses collective action from a conceptual point of view and also considers the structural motivational aspects inherent in situations of collective action. Collective actions are studied especially by social psychologists, sociologists, and economists as well as philosophers and researchers in computer science. Below the central conceptual features of collective actions are focused on and some relevant classifications are made.

The social connectedness of participants in a collective social action can, in its most central sense, be taken to refer to the social attitude, especially “we-attitude,” shared by the participants. The attitude can be a want, goal, intention, belief, wish, hope, feeling, and so on. A person’s we-goal can be schematically explicated as follows: A participant is said to have a we-goal \( G \) if and only if (a) he or she has the goal \( G \) (partly) because (b) he or she believes that the other participants have the goal \( G \), and (c) he or she also believes that it is a mutual belief among the participants that they have \( G \). The *genuineness* condition (a) is obvious. The *conformity* condition (b) serves to connect the participants subjectively, and the *social awareness* condition (c) connects them intersubjectively (if the belief is true). In the case of intentional action, the *because* will typically express a reason, in other cases a mere cause. For instance, people may contribute to a collective good (here, a goal) where the (expected) goal satisfaction will be a shared reason for action. Note that the social awareness of kind (c) is needed to exclude nonsocial cases such as Max Weber’s of people in the street simultaneously opening their umbrellas when it starts to rain. Also, fads and fashions clearly require mutual awareness. Note still that there is also weak collective action with a shared we-attitude where the *because* is replaced by *and*.

Let us consider some subcases of collective action performed because of a shared we-attitude. We concentrate on intentional cases (and ignore, e.g., invisible hand cases) and those in which the agents are
in a symmetric position. Our classification consists of four main classes of collective action constituted, respectively, by (1) independent participant action, (2) interdependent participant action, and (3) joint action. Also (4) group action, especially action performed by an organized group (e.g., business company) and attributed to a whole group—rather than to some persons collectively—is also a kind of collective action. (1) and (2) consist of actions based on the participants’ shared personal interests (“I-mode” interests) whereas at least in the full-blown cases of (3) group interests (“we-mode” interests) will serve as the participants’ reasons for action, such as Tom’s and Jane’s joint action of cleaning the house. With some exceptions, analytical philosophers have concentrated on (3) and cases with a small number of participants closely intentionally connected (in contrast to what we have in the case of large groups).

Category (1) typically contains cases such as people going for lunch to the same fashionable place for a social reason, typically a shared we-attitude. Category (2) contains cases involving strategic action in the sense of game theory. Here the belief about mutual dependency is central because rational action requires optimal response to the others’ actions. For example, people walking in a narrow corridor make an effort to avoid bumping into each other. Category (3) involves I-mode cases of joint action based on the participants’ shared personal intentions and beliefs. For instance, some people may jointly tidy up a road after a storm or celebrate a sports victory in the streets of a town. Category (3) also includes we-mode cases, in which the participants act as a group (e.g., lift a table or sing a song as a group). They are thus performing their individual actions for a group reason and jointly intend to perform their action together and, because of the joint intention, are collectively committed to its performance in a sense resulting in their “being in the same boat.” This kind of full-blown joint action must involve a joint intention to act together and appropriate shared beliefs making the participants interdependent. As to proper group action (category [4]), a group (e.g., a business company or a university) performs an action via its members (or a suitable subset by them, the “operatives”) performing a relevant joint action in virtue of which an action can be attributed to the group.

There are collective action dilemmas that involve a conflict between individual and collective rationality. Such dilemmas have in recent decades come to be regarded as central theoretical and practical problems in the social sciences.

While a collective action dilemma in its full sense is a many-person problem, this entry will concentrate on a simple two-person two-choice Prisoner’s Dilemma. It is given by the following matrix, where the rows represent the first person’s (or row player’s) action strategies or choices—“cooperate” (C) and “defect” (D)—and the columns the second person’s (or column player’s) similar choices:

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3,3</td>
<td>1,4</td>
</tr>
<tr>
<td>D</td>
<td>4,1</td>
<td>2,2</td>
</tr>
</tbody>
</table>

The first number in each cell represents the payoff (money or “utility”) to the row player, while the second number represents the payoff to the column player. There are four possible joint outcomes: CC, CD, DC, and DD. For example, the outcome CD is obtained when the row player chooses C and the column player chooses D. CD gives the payoff of one unit to the row player and four units to the column player. Note that the choice of C by the first player can lead either to CC or CD, depending on whether the second player chooses C or D. The first player’s preference ranking of the joint outcomes, as represented by the payoff numbers, is DC > CC > DD > CD (with “>” meaning “is preferred to”), and symmetrically for the second player. DC involves the row player’s “free riding” (failure to choose C), and it gives the highest payoff to him or her but the lowest (or “sucker”) payoff to the column player.

In this single-shot game situation, CC is preferred by both players to mutual defection DD, but DD is the only equilibrium—that is, an outcome in which neither player can rationally change his or her choice as long as the other one sticks to his or her choice. CC represents what collective rationality entails: The group formed by the players gains most (i.e., six units) by ending up in this joint outcome. In contrast, DD is taken to be the individually rational result that maximizes an individual’s expected payoff or his or her most secure choice. It is typically assumed here that the players do not communicate with each other in a way creating binding commitments.

Yet, even this simple case involves some central elements of a collective action dilemma of which the following account can be given: Letting S be a
game structure of the above kind, we can define that S involves a collective action dilemma if and only if S involves a conflict between collectively rational and individually (“privately”) rational action, that is, if and only if there are (distinct and incompatible) actions or strategies C and D such that C is a collectively rational, “cooperative” action and D is an individually rational action. This characterization fits (e.g., the Prisoner’s Dilemma and the game of Chicken) a game with the preference ranking DC > CC > CD > DD in the case of the row player. There is a free rider incentive—a participant individually gains by refraining to “cooperate,” as long as sufficiently many others participate, as required for the production of the good in question. The characterization also includes “take some” dilemmas such as the tragedy of the commons (e.g., pollution, fish depletion, and in general terms, refraining from excessive utilization of a resource) and “give some” dilemmas (e.g., national defense, voting, and in general doing one’s active share in the provision of a good, be it one in an institutionalized context or not). Consider air pollution. By reducing our individual environmental pollution by a small amount (this is the C choice) a remarkable improvement in air quality is achieved. There is a collective gain, which improves each individual’s situation relative to what universal defection (everyone’s choice of D) would yield.

These kinds of situations normally have the structure of either the Prisoner’s Dilemma or Chicken. However, there are also other kinds of collective action problems (e.g., coordination situations) with uncertainty of others’ contribution even if contributing is collectively rational (e.g., the Assurance game, defined by the preference ranking CC > DC > DD > CD, fits in here). These kinds of cases need not involve a free rider incentive.

Mancur Olson is famous for defending a number of claims about public goods and their provision, especially that large groups can at best provide suboptimal levels of public goods. The general conclusion about his theses is that they at least need special qualifications to hold true.

There is much both theoretical and experimental research concerning how people behave or tend to behave in collective action dilemmas, but here this research cannot be discussed except for a brief comment on a group-based (or we-mode) approach that contrasts with the prevalent individualistic (or I-mode) accounts. The we-mode approach explains why people often cooperate in dilemma cases where, for instance, rational choice theory recommends defection. Consider first what standard rational choice theory entails for our earlier single-shot Prisoner’s Dilemma example. If the agent—the first or “row” player, here “I”—thinks strategically and intends to maximize his or her value or utility, he or she can reason thus: I prefer the joint outcome DC to all the other joint outcomes; however, I realize that if you are reasoning similarly and planning to go for CD, we will end up in DD. Still, wanting to avoid the worst outcome CD, I cannot rationally aim at the collectively optimal CC. So I choose D and rationally expect DD to result in the single-shot case. In this case, my reason for choosing D is to secure at least the third-best alternative, given my beliefs and the other player’s rationality.

In contrast to the above case with individualistic or I-mode preferences, acting for a we-mode reason can lead to the group members’ (collectively) rationally cooperating in collective action dilemmas. Here the members adopt the group’s point of view and are disposed to act as interdependent and collectively committed group members. They (“our group,” “we”) accept to act on the higher payoff as a group and thus come to intend to choose C (over D). Thus, they cooperate as a group and choose C here as their part of the group’s achieving the group goal (typically a collective or at least a nonexcludable good in the group). The original dilemma dissolves.

The we-mode approach has also received experimental support. It applies especially to dilemmas where there is a collectively preferred joint outcome and the participants are or take themselves to be in the same boat with collective commitment. In institutionalized cases where the state in a quasi we-mode fashion provides public goods for the citizens (typically on the basis of taxes), strong sanctioning is often possible and even “I-moders” can be brought to act according to the norms issued by the state—in many cases partly against their will.

Raimo Tuomela

See also Decision Theory, Philosophical Perspectives; Joint or Collective Intention; Philosophy of Action

Further Readings

COMMON CODING

A common coding approach to perception and action was first proposed by Wolfgang Prinz and later elaborated by Bernhard Hommel and colleagues. The concept of common coding refers to the representational organization of interfaces between perception and action. Unlike traditional approaches in psychology and neuroscience, which posit separate and incommensurate representations for perception and action, common coding invokes that perception and action control draw on shared and commensurate representations. This entry discusses basic claims and major research issues associated with common coding as well as relationships to similar approaches.

Basic Claims

Commensurate Representation

Most classical frameworks of interfaces between perception and action rely on separate coding. The classical view holds that representations on the afferent and the efferent side of the interface are incommensurate. While afferent representations refer to patterns of sensory stimulation, efferent representations refer to patterns of motor activation. Bridging the gap between them requires creating arbitrary linkages (mapping between sensory and motor codes).

In contrast, the common coding approach suggests that in addition to such separate codes, there are further levels of representation in which afferent and efferent information have the same format and dimensionality. Common coding is meant to apply to the relationship between representations on the afferent and the efferent side (referring to ongoing events and intended actions, respectively). These representations are commensurate, since they both refer to external events. They permit creating linkages between perception and action that do not rely on arbitrary mappings. In common coding, action planning is conceived of in terms of operations that move from current events to intended future events. These operations involve the matching of event codes and action codes; perception and action may modulate each other by means of similarity.

Ideomotor Principle

The principle of common coding is closely related to the ideomotor theory of voluntary action as advanced by Rudolph Hermann Lotze and William James in the 19th century. In line with this theory, common coding hypothesizes that actions are represented in terms of their perceptual consequences. The representations of actions are thus similar to those of other events, except that they can be generated through bodily movements. When individuals perform actions, they learn what their movements lead to (ideomotor learning). Ideomotor theory claims that these associations can also be used in the reverse order: When individuals perceive events of which they know (from previous learning) that they may result from certain movements, perception of these events may evoke the movements leading to them (ideomotor control). The distinction between such learning and control is analogous to the distinction between forward and inverse computation in motor learning and control. While ideomotor learning leads to the prediction of action outcomes, ideomotor control helps in the selection of actions that will lead to an intended outcome.

Research Issues

Content-Based Interference

A straightforward application of common coding is to compatibility effects in task performance. Compatibility effects indicate that task performance
is modulated by similarity, or compatibility, between stimulus and response sets. For example, as has been shown by Marcel Brass and colleagues, common coding can account for movement imitation in stimulus-response compatibility paradigms. Other studies have extended the scope of the notion of compatibility to action effects. Hommel showed that stimulus-effect compatibility may under certain conditions override stimulus-response compatibility. Likewise, Wilfried Kunde showed that response-effect compatibility may be functional as well. These effect-related compatibility effects provide direct support for the claim that action selection engages anticipatory representations of action effects.

In a series of studies on visual action recognition, Günther Knoblich and colleagues have shown that observers can tell self-produced from other-produced action, indicating that visual action perception may engage resources for action production. These studies, too, make a strong case for shared representational resources for perception and action.

Content-based interference between perception and action has also been demonstrated in a number of dual-task settings. Jochen Müsseler and Hommel reported an inverted compatibility effect between response processing in one task and stimulus processing in a concurrent task (i.e., better performance on incompatible than on compatible trials). Gijsbert Stoet and Hommel reported similar interference effects between stimulus and response processing across two tasks that were nested in each other. A number of further studies from Hommel’s and Prinz’s labs demonstrated both regular and inverted compatibility effects between motion perception and movement production, making a strong case for shared representational resources for perception and action.

Movement and Goals

Ideomotor theory claims that action production implicates representations of action goals. Evidence in support of this claim comes from studies by Prinz, Harold Bekkering, Franz Mechsner, and Andreas Wohlschläger. They refer to such diverse research domains such as imitation, bimanual coordination, ideomotor action, and tool use action. Further studies by Gisa Aschersleben and Moritz Daum have underscored the role of action goals for action production in early infancy.

Interactive Coordination

In studies by Bekkering, Knoblich, Prinz, Natalie Sebanz and Wohlschläger the notion of common coding has also been applied to coordination of perception and action across individuals. Studies on imitation and ideomotor action provide ample evidence showing that individuals who watch others acting may become induced to act like them—either in terms of movements or goals. Further studies have addressed functional properties of action simulation, which is often seen to mediate the transition from action perception to action production.

Interactive coordination may also apply to task sets for coordinating perception and action. Coordination of task sets across individuals has been demonstrated in paradigms in which two individuals share a common task in a division-of-labor-mode. These studies suggest that individual performance does not only depend on one’s own share of the task but on the other’s share as well, suggesting corepresentation of the other’s task set.

Related Approaches

While most traditional approaches tend to stress the relative independence of perception and action, some theories have argued for closer links. For instance, Alvin Liberman and Paolo Viviani have proposed so-called motor theories that make a strong case for motor contributions to speech and action perception. The same applies to the theory of dimensional overlap advanced by Sylvan Kornblum. Further, following the lead of James J. Gibson, close representational connections between perception and action (albeit of a nonrepresentational kinds) have been claimed by various kinds of ecological approaches to perception, action, and cognition.
Today common coding theory is closely related to research and theory in two intersecting fields of study: mirror systems and embodied cognition. The notion of mirror systems is advocated by Giacomo Rizzolatti and colleagues. As concerns these systems, the principle of common coding seems to reflect major functional features that are implemented in mirror neurons and mechanisms in the brain. The notion of embodied cognition is advocated by Lawrence Barsalou and Arthur Glenberg. The notion of common coding is compatible with the claim that semantic meaning is embodied—that is, grounded in perception and action and the underlying sensory and motor operations.

Wolfgang Prinz

See also Attention and Action; Consciousness and Embodiment; Mirror Neurons

Further Readings


compound, the semantic transparency of the compound's head (e.g., the similarity of the meanings of ball and snowball) contributes more to the ease of processing than the transparency of the compound's modifier (snow in snowball), though the strength of this effect varies cross-linguistically with the position of the head in the compound (e.g., final in English and Dutch, initial or final in Basque and French). On a finer-grained semantic level, the relation between the compound's head and modifier (snowball, relation: MADE OF) affects how compounds are produced and the ease of recognition and learning of novel compounds for both adults and children.

Nevertheless, the notion that semantics is predominant in compound processing is currently being challenged by a series of experiments that demonstrate that the meanings of constituents are activated via computation in both transparent and opaque compounds. Also, there is evidence that even low-frequency and clearly transparent compounds leave traces in the mental lexicon and activate the storage route. Studies using behavioral measures with fine temporal resolution confirm that semantic effects only emerge at late stages of lexical processing and that the storage versus computation biases are primarily sensitive to properties of compounds' form rather than meaning.

Form and Analogy

Recent cross-linguistic research has shown that compound processing implicates activation both of compounds as wholes and of their constituents. Linguistic properties of compounds and constituents (e.g., the frequency of occurrence or orthographic length) demonstrably affect acoustic characteristics of the production of compounds in speech, as well as the time it takes listeners or readers to recognize a compound during auditory or visual comprehension. Moreover, morphemes deeply embedded in larger constituents of compounds (-er in dishwasher) influence lexical processing, suggesting that morphological parsing reaches lower hierarchical levels of word structure. The order of accessing constituents in opaque as well as transparent compounds appears to be left-to-right in production and comprehension, as if the constituents were presented distinctly as regular words in a sentence.

An important body of work has identified a range of cues that facilitates segmentation of constituents out of compounds, favoring computation. These include salient markers of the boundary between compound constituents: hyphens, capitalized letters or spaces, and low-frequency combinations of characters or phonemes straddling the boundary (e.g., wb in snowball). Additional work has shown that short compounds are biased toward the storage route, while long compounds favor computation. Taken together, the findings corroborate the current models of morphological processing that advocate the parallel use of storage and computation during processing. These models also embrace Gary Libben’s notion of the “hungry lexicon,” which suggests that rapid and accurate lexical recognition and production of compounds hinges on the simultaneous and interactive processing of multiple sources of information. In other words, the relevant information is processed as soon as it becomes available at respective levels of sound, form, or meaning.

Robust cross-linguistic findings demonstrate that relations between semantically or formally similar groups of compounds (morphological paradigms or families) are also important determiners of the lexical representation and processing of compounds. Cross-linguistic experiments on existing and novel compounds with both adult and child participants establish that compounds with constituents that occur in a large number of other compounds (e.g., fireman, policeman, postman) tend to be processed faster and more accurately than compounds with constituents that have less paradigmatic support. The morphological paradigm of the compound’s constituents has also influenced the choice, placement, and acoustic realization of the linking element in Dutch and German compounds (-s- in sportsman). These observations provide support for the view of morphology as probabilistic, rather than deterministic, regularities in mappings between word forms and meanings and shed light on the architecture of the mental lexicon underlying compound processing.

Future Directions

Over the past 40 years, the processing of compound words has proved an insightful area of psycholinguistic exploration, and research on this topic continues to expand over languages with different morphological structure, testing how different populations of speakers make use of their mental lexicon in real-time language production and comprehension tasks.
Prospective research topics include the refinement of theoretical models of morphological processing, the acquisition of compounding by children and L2 learners, and the neurophysiological foundations of unification of constituents into a compound meaning.

Victor Kuperman

See also Conceptual Combination; Word Learning; Word Recognition, Auditory; Word Recognition, Visual

Further Readings


**Computational Models of Emotion**

A computational model of emotion describes structures and processes related to emotion, mood, feeling, affect, or other related phenomena in sufficient detail such that it can be implemented on a computer. There are at least three kinds of computational emotion models: those that recognize emotion (e.g., recognize facial expressions), those that express emotion (e.g., generate predetermined facial expressions), and those that model human emotion processing (such as simulate how the inability to achieve a goal may lead to frustration). For models that simulate emotion, there are both high-level models that abstract away from details of the brain and attempt to describe all the interacting components involved in emotion, and low-level models that focus on the details of specific brain systems.

A model that recognizes emotion can be useful in improving human-computer interaction by providing additional information about a human user to the computer. For example, if a tutoring system can detect that a student is frustrated or confused, it can dynamically alter its teaching strategy to help the student. Sources of information that can be used to inform an emotion recognition system include facial expressions, vocal qualities, physiological measures (skin conductance, heart rate, blood pressure, etc.), and brain imaging.

A model that expresses emotion, usually through facial expressions or voice modulation, can improve interactions with humans by providing an indirect way of indicating the state of a computer system—whether it is being successful in its task or not and whether the human using the system is making progress on a task. Emotion expression also can enhance computational entertainment systems by making the behavior of virtual characters more realistic.

The remainder of this entry focuses on models described as *simulating* emotional processing, Rosalind Picard lists five criteria that must be met for a model to fully simulate emotion. In her words, such a system *has* emotion. In this article, we will finesse the debate over the differences between simulating and having emotion and concentrate on whether existing models achieve these criteria or not. Picard’s five criteria for a model are as follows:

1. It engages in emotional behavior; some of its behavior is a consequence of its emotions.
2. It supports fast primary emotions, which are automatic emotional responses that can occur with little or no cognitive input.
3. It supports cognitively generated emotions, which are emotions that arise from more deliberate processing.
4. It has an emotional experience; it has subjective feelings.
5. There are body-mind interactions; emotions have an influence on physiological and cognitive processes and vice versa.

A system that achieves only a subset of these criteria can still be useful.

The criteria imply the need for not just a model of emotion but also a complete model of cognition. Indeed, most systems that claim to be nearly complete are implemented in cognitive architectures. Cognitive architectures are typically composed of a small number of interacting computational mechanisms (memories, processors, and interfaces) that give rise to complex behavior. Some architectures have been designed around emotion while others have been augmented with emotional mechanisms. Given the broad scope of behavior these systems
attempt to model, cognitive architectures tend to be
implemented at the symbolic, as opposed to neural,
level (although hybrids exist).

Most existing computational models of systems
that simulate emotion are based on an appraisal
theory. Appraisal theories postulate that a situation
is evaluated along a set of fixed dimensions, many
of which relate the situation to the agent’s goals,
such as goal relevance, goal conduciveness, agency,
novelty, and so on. They usually (but not always)
describe emotions as categories, postulating a dozen
or so different emotions (joy, fear, anger, guilt, etc.).
These theories are used to explain how emotion
arises from cognitive processing and how an agent
can cope with its emotions by either internal actions,
such as changing goals or its interpretation of the
situations, or external actions, which attempt to
change the situation. Since these theories are cogni-
tively oriented, appraisal theories fit naturally within
the context of cognitive architectures.

The strength of these models is their breadth
and integration of the different aspects of emotion
with the rest of cognition, providing complete theo-
ries of how emotion is produced and how it influ-
ences behavior and learning. These models make
predictions that can be compared to overt human
behavior, such as self-reports of emotions in real or
imagined situations. In addition to psychological
modeling, emotion has proved valuable in increasing
the functionality of cognitive architectures through
improving decision making, planning, and learning.
The weakness of these models is that they abstract
away from many details of human behavior, making
it difficult to validate them using physiological data,
such as that available from brain imaging.

Neural network models are at the other end of
the spectrum, focusing on brain structures known to
be involved in emotions (e.g., the amygdala and lim-
bic system) and how they interact with higher level
cognition and attention (dorsolateral prefrontal cor-
tex) and other areas. Some models include computa-
tional theories of the interactions between emotion
systems and learning (e.g., via nucleus accumbens).

Neural network models tend to adopt a circum-
plex approach to emotion, in which emotions are
described along two or three dimensions, such as
arousal and valence. Superficially, this is similar to
appraisal theories; however, here there are fewer
dimensions and they describe properties of the emo-
tion, not of the relationship between the situation
and the agent. Models of specific emotional processes
(e.g., fear conditioning) and mental disorders (e.g.,
depression) also exist.

The strength of neural network models lies in
their ability to model directly measurable phe-
nomena, and hence they can be directly validated.
The weakness of these models is that the low-level
interactions of the various parts of the brain are not
sufficiently understood to develop large-scale mod-
els. Furthermore, the computational requirements
of simulating neural networks restrict them in size,
making building complete models impractical.

Future work will likely see these approaches
merge, as high-level models attempt to explain lower
level phenomena and low-level approaches attempt
to broaden. An area in need of exploration from
both sides is embodied emotions, in which a realistic
model of the body is integrated with models of the
brain. Many phenomena, from hormone levels to
action tendencies, will likely remain poorly under-
stood until this leap is made.

Robert P. Marinier III and John E. Laird

See also Attention and Emotions, Computational
Perspectives; Decision Making and Reward,
Computational Perspectives; Emotion,
Psychophysiology of; Emotion, Structural Approaches;
Facial Expressions, Computational Perspectives

Further Readings


structure of emotions. Cambridge, UK: Cambridge
University Press.

MIT Press.

CONCEPTS, COMPARATIVE
PERSPECTIVES

The question of what it means to have a concept
becomes even more difficult to address in the com-
parative literature—that is, in research studying con-
cept formation across a variety of species. Typically,
to form and hold concepts means to have an internal
representation of an object, event, or construct that
allows one to categorize exemplars, or members of the category, in a meaningful way. This ability to represent the world in terms of concepts means that one can respond to novel exemplars one encounters in appropriate and adaptive ways. For instance, a child does not need to learn that every individual vehicle could be dangerous if stepped out in front of when crossing the street. Rather, once the child learns that any large, fast-moving vehicle could be harmful, and can categorize trains, streetcars, buses, trucks and cars of various types into the “vehicle” category, the child can respond in a safe and appropriate manner when he or she encounters one in his or her environment. That is, once the child has a concept of *vehicle* and what vehicles are and what they can do, that child can respond to all members of that category in an appropriate fashion. This entry will address the question of what sorts of concepts nonhumans can form, if any.

**Do Animals Have Concepts?**

It is a longstanding empirical question in the comparative literature whether animals similarly form concepts. Richard Herrnstein and colleagues pioneered some of the earliest studies with pigeons when they asked whether the birds could learn to discriminate between pictures of humans and pictures that did not include humans—thus indicating that they had a concept of *human*. The researchers then went on to show that pigeons also had concepts for water and trees and other natural objects. The early concept formation literature was thus dominated by studies aiming to show that animals formed many natural concepts, at many levels of abstraction. Following in the important footsteps of Eleanor Rosch’s seminal work with human children, it was important to show that animals could form abstract as well as concrete level categories. Very concrete categories—those in which the exemplars all share many physical features in common—can be discriminated solely on the basis of attending to perceptual features. If animals respond differently to pictures of humans versus nonhumans solely by looking for and responding to skin-colored patches or eyes or clothing, this would not indicate that they truly held a concept of *human*, but rather that they responded to the world as a conglomeration of perceptual features, not as conceptual wholes. However, if they could form categories for things that were not bound together by perceptual features, such as functional categories such as *foods* and *tools* and *toys* and *animals*—categories that can include exemplars that are perceptually quite diverse—or even more abstract categories that include nonobservable entities such as “things that make one happy”—this would constitute evidence that animals also form concepts.

**Methods of Study**

Thus, comparative researchers interested in the general topic of whether animals form concepts have focused on diverse but convergent questions. Some have continued in the tradition of the early research, presenting animals with various pictorial stimuli and asking them to categorize the stimuli, using two-choice discrimination paradigms, in which they select images that belong to one category, and not images that belong to another, or simply that don’t belong. Alternatively, they might use a match-to-sample paradigm, in which animals select from comparison images the one that matches the category depicted in a sample image. Experimenters can then rigorously manipulate and examine the stimuli chosen and rejected by their subjects in order to determine what features are being used to drive their choices. Recently experimenters have devised new techniques, such as bubbles, for isolating and illuminating salient features of the stimuli to present and measuring the impact on the animals’ choices. Alternatively, researchers can now use more flexible software to directly manipulate the features in a stimulus to examine the individual effects of, for example, enlarged eyes, elongated legs, darkened hair, and so on. While it is difficult to adequately summarize the outcome of the various studies across all of the species studied, suffice to say that most animals use a variety of salient perceptual features to aid them in performing these categorical discrimination tasks—some of which are categorically relevant, such as the presence of eyes when discriminating animals from nonanimals, and some not categorically relevant, such as preferring close-up to distant images when discriminating humans from nonhumans. Thus, this large body of literature remains agnostic on the question of whether animals truly form concepts in the same manner that humans do, independent of specific individual features but is suggestive that animals are much more perceptually biased rather than forming overarching abstract concepts.
Concepts for Unobservables

Another prolific body of research has focused on the equally fascinating question of whether nonhumans are capable of forming concepts for unobservables—things that cannot be directly perceived through the senses. One particular class of unobservable constructs includes mental states, such as thoughts, beliefs, and feelings. Ever since David Premack and Guy Woodruff posed the famous question, “Does the chimpanzee have a theory of mind?” researchers have wondered if nonhumans reflect on the thoughts of others. These types of questions are particularly challenging for scientists as it is difficult to determine when one is predicting another’s future actions on the basis of their underlying thoughts or beliefs or simply on their past actions or contextual cues. Recently, a large corpus of data has suggested that chimpanzees at least do possess a theory of mind; that is, they are capable of forming concepts for some mental states, such as seeing, intentions, and perhaps even knowing. For instance, they respond differently to humans, and other chimpanzees, depending on what they have seen, and therefore what they know, about things such as where food might be hidden in competitive foraging situations. However, these findings are not without an alternative explanation. Thus, the question of what sorts of concepts nonhumans hold is still an open empirical question—one that is as intriguing as it was when Darwin first pondered, is the difference between man and other creatures one of degree or one of kind?

Jennifer Vonk

See also Categorization, Neural Basis; Categorization, Psychological Perspectives; Classical Theory of Concepts; Concepts, Development of; Concepts, Philosophical Issues; Concepts and Language

Further Readings


CONCEPTS, DEVELOPMENT OF

Concepts are critically important to a range of human activities and, accordingly, are widely varying in content, process, structure, and function. There are competing theoretical accounts for how concepts develop, ranging from the position that there are qualitative changes in concepts with age to the view that foundational concepts are innate. However, despite this debate, there is widespread agreement for early conceptual understanding even in infancy. Children’s concepts have broad implications for conceptual development more broadly, including stereotyping, knowledge acquisition, and cultural influences on thought.

Concepts are mental representations that organize experience. They are central to the full range of psychological activities, including recognition (an infant smiling on viewing a face), language use (a toddler requesting a snack), problem solving (using analogy to figure out the structure of benzene or the orbits of planets), and theory construction (Darwin’s introduction of natural selection). They provide an efficient means of storing information (e.g., rather than remembering every single apple one has encountered in the past, one can instead update a general representation of apples), permit making inferences into the future (e.g., inferring that a newly encountered apple is edible), and form the building blocks of ideas (e.g., the thought “Apples are juicy” requires the constituent concepts apples and juicy). Within psychology, the most-studied human concepts are categories of physical objects (apple, horse, table), but our concepts also include properties (happy), events or states (running),
individuals (Fido), and abstract ideas (fairness). One of the hallmarks of human cognition is our ability to impose order on experience and organize it flexibly according to our needs.

Scholars have studied children’s concepts as a means to investigate classic developmental issues, including the question of whether there are innate concepts, continuity, and change in thought; the domain generality versus specificity of human cognition; and the origins of complex reasoning. Studies of concept development also provide insights on children’s knowledge and beliefs about the world. For example, much research has investigated children’s concepts of particular core domains of thought such as mental states (“theory of mind”), biological processes, physical entities, number, space, and time.

**Different Kinds of Concepts**

Children acquire a wealth of different kinds of concepts, and these different concepts develop in different ways. For example, how students acquire a mathematical concept in a classroom may differ in important ways from how infants acquire a concept of animacy. Concepts vary from one another in content (mathematical, biological), process (learned explicitly in school vs. implicitly in ordinary interactions), structure (taxonomic relatedness vs. thematic relatedness), and function (allowing quick identification of prey while hunting vs. determining scientific classification). Furthermore, certain concepts are important for influencing other concepts and, thus, can be considered foundational. A concept such as cause, for example, implicitly guides children’s learning; a concept such as alive has implications for how one thinks about animals, plants, and self-moving artifacts.

**Concepts and Language**

An enduring question is the relation of language to concepts. Research with infants, as well as with nonhuman species, demonstrates conclusively that organisms do not require language in order to learn and use complex and sophisticated concepts. Preverbal infants have rich expectations regarding number, causation, animals, containment, physical support, and so forth. At the same time, language appears to emphasize and support certain types of reasoning. Infants and young children treat words as “invitations” to form a category and search for commonalities. By 9 months of age, children more readily notice similarities among objects that receive the same name, compared to objects that receive different names or that receive no labels at all. Likewise, words encourage children to extend their knowledge in new ways. For example, if a preschool-aged child learns a new fact about one bird, she will infer that other birds have that same property, even other birds that look very different from the bird they first learned about. Moreover, the form of language influences the kind of expectations children reach, with nouns more than adjectives leading children to treat categories as stable and inference promoting. For example, children think that a person labeled with a noun (“He is a carrot eater”) is more likely to eat carrots his whole life than is a person described with a verbal predicate (“He eats carrots whenever he can”).

**What Develops? Contrasting Developmental Theories**

Historically, children’s concepts were characterized in terms of what they lacked: Whatever it was that adults have, children were thought not to have. Thus, classic developmental theorists such as Jean Piaget and Lev Vygotsky proposed a host of “developmental dichotomies”: Children are perceptual, adults are conceptual; children are thematic, adults are taxonomic; children are concrete, adults are abstract. However, these dichotomies have been undermined by compelling results with young children making use of subtle experimental techniques. Rather than asking children to verbally articulate their concepts, these methods rely on more implicit measures (how long an infant looks at an image; the sequence in which an array of objects are touched; inferences that require only simple pointing judgments). When tested in these ways, preschool children and even infants consider an impressive range of cues, features, and conceptual structures. Early concepts are not wholly perceptual, as children reason about intentions, number, and causation. Early concepts are not wholly thematic, as children detect commonalities among items that rarely appear together in the real world (e.g., kangaroos and polar bears are both animals). Early concepts are not wholly concrete, as children have the capacity to talk and reason about abstract “kinds” of
things (e.g., birds as a general category). Indeed, it is not even clear that concrete representations are helpful to children, as at times they stand in the way of children’s understanding of more abstract concepts (such as letters of the alphabet or arithmetic).

Given children’s impressive capacities to learn and reason, an important question is how children acquire such a rich set of concepts in such a short amount of time. There are at least three mechanisms that researchers have identified, all of which have some support (though scholars debate the relative role of each): innateness, associative learning, and theory construction. The idea that there are innate concepts is controversial but supported by the finding that even newborns appear to have consistent expectations in particular domains. For example, infants expect physical objects to maintain a continuous, connected path over time, to maintain a connected, bounded whole (e.g., not to split apart into two separate objects), and to bypass other objects when moving through space pass (i.e., not to pass through other objects), suggesting that certain assumptions about the properties of physical objects have innate support. Other concepts are readily acquired on the basis of associative learning. Human infants possess powerful mechanisms for detecting regularities in the environment and learning associations. As do other organisms, child learners who experience two properties in close and frequent co-occurrence will learn to associate those properties together. Infants readily acquire associations between properties such as shape, movement, and texture and make use of such associations to learn some of the differences between animals and artifacts. Still other concepts appear to require causal reasoning and integration of facts within a growing commonsense “theory.” For example, children’s understanding of what living things have in common undergoes marked change as children gain a deeper understanding of the causal principles involved in biology.

Implications for Conceptual Development

The ease with which young children acquire concepts has broad implications for other aspects of childhood thought, including stereotyping, knowledge acquisition, and universality/variability across cultures.

Stereotyping

In their early developing concepts, children at times overlook the variability within categories, focusing instead on the differences between groups. This tendency may encourage the acquisition of social stereotypes. For example, 4-year-olds infer that two girls will engage in similar behaviors, even if one of them looks more like a boy or has previously chosen to engage in male-stereotypic behaviors. Young children’s gender concepts, more so than those of adults, emphasize the differences between groups. Preschoolers view gender stereotypic properties (ballet for girls or football for boys) as automatically acquired, whereas adults acknowledge individual variation in the development of these properties. Moreover, young children expect gender differences to be deeply reflective of other properties, including a range of physical, behavioral, and internal differences.

Knowledge Acquisition

One of the primary functions of concepts is to enable people to make inferences about future events. When children learn a new fact about one category member (for example, that a sparrow has hollow bones) they readily generalize this new knowledge to other category members (such as other birds). This way, children’s concepts facilitate knowledge acquisition. Concepts can also interfere with learning, however, by making it difficult to understand new information that conflicts with an individual’s current understanding. For example, when children focus on within-category similarities (e.g., the properties shared by all birds), they often have difficulty understanding the process of evolution by natural selection, which rests on an appreciation of the meaningful variability that exists within categories. Similarly, intuitive concepts of physical objects, which include expectations about object solidity, may interfere with learning more advanced theories of physics. Confronting errors in children’s existing concepts is therefore important for developing successful educational methods.

Universality and Cultural Variability

Children’s concepts reflect both universal tendencies and culture-specific variation. Research conducted in a range of diverse cultures supports the conclusion that there are certain cognitive biases that universally guide conceptual development. For example, young children across diverse cultural contexts assume that categories of living things (wolves, fish, trees) capture an underlying reality. They believe
that membership in such categories (e.g., being a wolf), as well as associated properties (e.g., having sharp teeth), are determined by a deep, intrinsic, and stable property that is acquired through natural processes (e.g., inheritance from wolf parents) (termed psychological essentialism). There is also evidence, however, that cultural input influences conceptual development across childhood. For example, cultural contexts influence how children make sense of human behavior. Children in relatively more collectivist communities (as in China) are more likely to view behavior as determined by an individual's context, whereas children in relatively more individualistic communities (as in the United States) are more likely to view behavior as resulting from individual factors such as personality traits.

Conclusions

Children's concepts are remarkably sophisticated, from infancy onward, and reveal a diversity of ways of viewing the world. Children's concepts also provide important tools for further learning. At the same time, concepts can lead children to biases and stereotypes, and their concepts undergo tremendous change throughout childhood. In these ways, children's concepts provide a microcosm of the wealth of theories that humans construct over development.

Susan A. Gelman and Marjorie Rhodes

See also Categorization, Psychological Perspectives; Concepts, Comparative Perspectives; Concepts and Language; Knowledge Acquisition in Development; Representations, Development of

Further Readings


Concepts, Philosophical Issues

In this entry, the main theories of and the current controversies about concepts in the philosophy of psychology are reviewed. The relation between philosophers’ and psychologists’ theories of concepts is first discussed. Then, the two principal kinds of philosophical theories of concepts are briefly introduced. Third, the four main types of psychological theories of concepts are presented. Then, the recent debate about concept pluralism and concept eliminativism is examined. Finally, the controversy about empiricism is surveyed.

Philosophers’ and Psychologists’ Theories of Concepts

The term concept is used in philosophy and in psychology, and philosophers, such as Georges Rey, Eric Margolis, and Jerry Fodor, commonly assume that psychologists’ theories of concepts aim at solving the issues they themselves are interested in. In addition, they commonly hold that as solutions to the issues of interest in philosophy, psychological theories of concepts are defective. For instance, in 2003, Fodor concludes his review of Gregory Murphy’s book, The Big Book of Concepts, as follows:

It is part of our not knowing how the mind works that we don’t know what concepts are or what it is to have one. Just about everything that current cognitive science says about either topic is wrong. But at least it is clear that concepts aren’t typicality structures and that having them is not being able to sort things. Except for leaving that out, Gregory Murphy’s book tells you most of what there is to the psychology of concepts. Read it, therefore, by all means; but don’t even consider believing it. (p. 4)
Philosophical theories of concepts are also sometimes criticized for being unable to explain what psychological theories of concepts are meant to explain—namely, how we categorize, make inductions, and so on.

The lack of progress in the resolution of the controversy between philosophers and psychologists suggests an alternative picture of the relation between philosophers’ and psychologists’ theories of concepts: These theories are not meant to solve the same issues and are thus not competing. In support of this alternative picture, one can note that philosophers and psychologists disagree about what phenomena a theory of concepts should explain. A meticulous examination of the psychological and the philosophical theories of concepts also shows that their explananda differ.

It is thus important to distinguish clearly the issues philosophers and psychologists are trying to solve:

1. **The philosophical issue**: How are we able to have propositional attitudes (beliefs, desires, etc.) about the objects of our attitudes? For example, in virtue of what can we have beliefs about dogs as such or about triangles as such?

2. **The psychological issue**: Why do people categorize, draw inductions, make analogies, combine concepts, and so on, the way they do? For instance, why are inductive judgments sensitive to similarity?

Philosophers attempt to solve the philosophical issue by providing sufficient (or necessary and sufficient) conditions for having propositional attitudes about the objects of our propositional attitudes. Psychologists attempt to solve the psychological issue by determining the properties of the bodies of information about categories, substances, events, and so forth, that people rely on when they categorize, make inductions, draw analogies, understand words, and so on.

It is thus clear that psychologists and philosophers do not attempt to answer the same questions. Psychologists attempt to explain the properties of our categorizations, inductions, and so forth, but they do not attempt to determine the conditions under which people are able to have propositional attitudes about the objects of their attitudes. By contrast, philosophers attempt to determine these conditions but not to explain the properties of our higher cognitive competences.

Taking this conclusion for granted, one could endorse two different views about the relation between the philosophical and the psychological theories of concepts. One could propose, as Susan Carey or Eric Margolis and Steve Laurence do, that these theories focus on different aspects of the same entities. Instead, one could propose, as Edouard Machery has done, that “concept” refers to different entities when used by psychologists and philosophers. In brief, Machery favors the second position because, if the first position were correct, then, when prototype theorists say that concepts are prototypes or when exemplar theorists say that concepts are sets of exemplars, what they say would be literally mistaken. The principle of charity suggests that those psychologists are better understood as theorizing about different things than philosophers.

**Two Kinds of Philosophical Theories of Concepts**

There are two main kinds of theories of concepts in philosophy: inferential and atomist theories of concepts. According to inferential theories of concepts, as defended, for example, by Ned Block, Christopher Peacocke, and Bob Brandom, to have a concept of x (viz., to be able to have propositional attitudes about x as such) involves our capacity to draw inferences. For some inferentialists such as Block, it involves being disposed to draw some inferences, while for others, such as Peacocke, it involves being disposed to take some inferences to be primitively justified—that is, to be merely justified in virtue of the concepts involved in the inferences. Logical concepts, such as the concept and, are usually taken to be the best illustration of the inferentialist theories of concepts. To possess and (to be able to think about the conjunction of states of affairs) is either to be disposed to draw the inferences represented by the and-introduction rule (that is, the rule allowing the inference of a conjunctive proposition connecting two propositions by means of and when one holds both propositions independently) and the and-elimination rule (that is, the rule allowing the inference of one of the two propositions connected by and in a conjunctive proposition when one holds this conjunctive proposition) or to be disposed to take these inferences to be primitively justified. Because inferences involve
several concepts, inferentialists hold that possessing a given concept always involves possessing other concepts.

Atomists hold that having a concept of \( x \) consists in standing in some relation with the property of being an \( x \). Particularly, Fodor holds that having a concept of \( x \) is to stand in a nomological relation (i.e., in a relation that has the status of a law of nature) with the property of being an \( x \) such that a mental symbol occurs (one entertains thoughts about \( x \)) when one perceives the instantiated property of being an \( x \). For instance, roughly, to have the concept cat is to be disposed to entertain thoughts about cats whenever a cat is present in one’s environment. On this view, one can have a single concept, whence the name “atomism.”

The Four Main Psychological Theories of Concepts

There are four main kinds of theories of concepts in psychology as discussed by Murphy in his 2002 book and Machery in his 2009 book: the classical theory of concepts, prototype theories of concepts, exemplar theories of concepts, and theory theories of concepts. Although there are several distinct prototype theories of concepts, these theories agree about the distinctive features of prototypes. The same is true of exemplar and theory theories.

According to the most common versions of the classical theory of concepts, a concept of \( x \) represents some properties as being separately necessary and jointly sufficient to be an \( x \). The concept of grandmother is perhaps the best illustration of this approach to concepts: If people have a classical concept of grandmother, they hold that to be a grandmother it is necessary and sufficient to be the mother of a parent. Most psychologists and philosophers of psychology have abandoned the classical theory of concepts for the reasons given by Murphy and by Machery in the Chapter 4 of his book. Anna Wierzbicka and Jacob Feldman are exceptions to this trend.

In substance, prototypes, as theorized, for example, by Eleanor Rosch and James Hampton, are bodies of statistical knowledge about a category, a substance, a type of event, and so on. For instance, a prototype of dogs could store some statistical knowledge about the properties that are typical or diagnostic of the class of cats. According to prototype theories, when one categorizes, draws an induction, makes an analogy, and so forth, one spontaneously brings to mind the properties that are typical, diagnostic, and so forth, of the relevant category, substance, and so on. Prototypes are typically assumed to be used in cognitive processes that compute the similarity between a prototype and other representations, such as the representations of the objects to be categorized, in a linear manner.

Exemplars, as theorized, for example, by Douglas Medin in 1978 and Robert Nosofsky in 1986, are bodies of knowledge about individual members of a category (e.g., kitty, tiger), particular samples of a substance, particular instances of a kind of event (e.g., my last visit to the physician). For instance, according to exemplar theories, a concept of cats would consist of a set of bodies of knowledge about specific cats (kitty, tiger). When one categorizes, draws an induction, makes an analogy, and so on, one spontaneously brings to mind the properties of specific members of the relevant categories, of specific samples of the relevant substances, and so forth. Exemplars are typically assumed to be used in cognitive processes that compute the similarity between a set of exemplars and other representations, such as the representations of the objects to be categorized, in a nonlinear manner.

Theories, as theorized by Murphy and Medin and by Carey, for example, are bodies of causal, functional, generic, and nomological knowledge about categories, substances, types of events, and so forth. A theory of cats would consist of some such knowledge about cats. When one categorizes, draws an induction, makes an analogy, and so forth, one spontaneously brings to mind this causal, functional, generic, and nomological knowledge. Recent work on causal knowledge by Alison Gopnik and colleagues suggests that theories might be used in cognitive processes that are similar to the algorithms involved in causal reasoning.

Concept Pluralism and Concept Eliminativism

What is the relation between the psychological theories of concepts introduced in the previous section? The received view is that these theories are competing: It is often assumed that concepts are either definitions or prototypes or sets of exemplars or theories. Recently, however, psychologists
and philosophers of psychology have proposed to look at these theories differently. Two positions have emerged: concept pluralism and concept eliminativism.

Concept eliminativists, such as Machery and Fernando Martinez-Manrique, hold that there are different kinds of concepts—particularly, prototypes, exemplars, and theories—and that the principal psychological theories of concepts describe them. In addition, concept eliminativists propose that these three kinds of concepts have little in common and that there is no point attempting to describe the properties of concepts that are common to all concepts. Finally, because these kinds of concepts are very different from one another and thus do not form a natural kind (i.e., roughly, a class of objects that share many properties in common and that is an appropriate target of scientific investigation), concept eliminativists argue that the notion of concept should be eliminated from the theoretical vocabulary of psychology.

Concept pluralists, such as Gualtiero Piccinini and Dan Weiskopf, agree with concept eliminativists that there are different types of concepts. The main difference between pluralists and eliminativists is that pluralists reject the proposal to eliminate the notion of concept, either on the grounds that there are many similarities between kinds of concepts or on the grounds that the notion of concept is useful to distinguish those animals that have concepts from those that do not.

**Neuro-Empiricism**

A number of cognitive scientists, such as Arthur Glenberg, Lawrence Barsalou, Jesse Prinz, and Alex Martin, have recently revived the empiricist approach to concepts. These neo-empiricist theories propose that entertaining a concept consists in entertaining some perceptual representations and that cognitive processing consists in manipulating such representations. Amodal theories deny these two claims.

For example, according to neo-empiricists, someone’s concept of pear consists of the visual, olfactory, tactile, somatosensory, and gustative representations of pears that are stored in his or her long-term memory. Retrieving the concept of pear from long-term memory during induction or categorization consists in producing these perceptual representations. This process is called “simulation” or “reenactment.” Cognitive processes involve manipulating the perceptual representations that are reenacted. By contrast, according to amodal theorists, someone’s concept of pear stores some perceptual (visual, tactile, gustatory, etc.) as well as some nonperceptual information about apples in a single, distinct, nonperceptual representational format. Cognitive processes are defined over these nonperceptual representations.

The notion of perceptual representation is in need of clarification. Barsalou has proposed that amodal representations are linguistic, while modal representations are analogical. This clarification is not without problems. Analogical representations are usually thought to be such that some properties of their vehicles co-vary with what is represented. Maps and mercury thermometers are good examples. While evidence shows that there are some analogical representations in the brain—retinocentric maps, for instance—there is no evidence that analogical representations, so understood, are pervasive in the brain. Moreover, because some representations are analogical, but do not belong to any perceptual system, as argued by Machery and by Guy Dove, it seems inadequate to contrast amodal and perceptual representations by means of the notion of analogical representations. Prinz has proposed that perceptual representations are whatever psychologists of perception say perception involves. This proposal is not without problems either. If psychologists of perception propose that perceptual representations are similar to traditional amodal symbols, for instance, if perceptual representations form linguistic representational systems, neo-empiricism would propose that our conceptual knowledge is stored in several linguistic systems. In this case, the distinction between neo-empiricism and the amodal approach would be rather thin.

Psychologists and cognitive neuroscientists have accumulated a wealth of experimental results to support the claim that concepts consist of perceptual representations. Philosophers have argued that the traditional objections against empiricist theories of concepts can be resisted.

By contrast, several researchers, such as Machery, Dove, and Bradford Mahon and Alfonso Caramazza, have identified various shortcomings of neo-empiricist theories and of the research supporting them. Machery has identified three problems with the neo-empiricist research. First, the empirical
results touted by Barsalou, Prinz, and others do not undermine the amodal theories of concepts in general. Rather, they provide evidence against specific amodal theories of concepts, while being naturally accommodated by other amodal theories of concepts (a problem called “Anderson’s problem”). Second, several experiments showing that people use perceptual representations fail to provide evidence against amodal theories of concepts because amodal theorists expect people to rely on imagery in these experiments (“the problem from imagery”). Finally, it could be that perceptual symbols constitute only a kind of concepts (“the generality problem”). In fact, research suggests that at least some conceptual representations—namely, the representations of the magnitudes of classes of objects or sequences of sounds—are amodal. Although these representations do not form a language, they are not perceptual either. Dove has developed the generality problem in great detail, showing that the research in support of neo-empiricism has typically focused on a single kind of concepts—namely, “concrete or highly imageable concepts”—and that neo-empirist findings are unlikely to be found in experiments involving concepts with low imageability such as abstract concepts.

Others have identified further difficulties. Reviewing a range of cognitive-neuroscientific work on concepts and various important behavioral studies, Mahon and Caramazza grant that the perceptual and motor systems are often activated during conceptual processing, but they insist that this activation falls short of supporting neo-empiricism. The reason is that this activation may be a simple by-product: The activation of these brain regions results from the activation of other brain areas, ones not involved in perceptual processing, and from this activation spreading from the latter areas to the former.

Finally, as reviewed by Martin in 2007, for example, many fMRI studies show that tasks meant to tap into the processes underlying higher cognition (particularly, tasks involving the understanding of particular words) activate either the very brain areas involved in perceptual and motor processing or brain areas near those. While neo-empiricists have often touted these findings as supporting neo-empiricism, others, such as Machery, hold that they actually challenge this approach. The reason is that since neo-empiricists insist that retrieving a concept from long-term memory consists in producing some perceptual representations, they are committed to the claim that the retrieval and use of conceptual knowledge activates the cortical areas used in perception and action. However, a typical finding in cognitive neuroscience is that the brain areas activated are near, and thus not identical to, the brain areas involved in perceptual or motor processing. Furthermore, these brain areas are almost always anterior to the brain areas activated in perceptual processing. A plausible interpretation is that the brain areas activated in the tasks tapping into higher cognition are amodal representations, which are distinct from the perceptual representations activated in the tasks tapping into perceptual processes.

Edouard Machery

See also Atomism About Concepts; Categorization, Psychological Perspectives; Classical Theory of Concepts; Concepts, Comparative Perspectives

Further Readings


**Concepts and Language**

Concepts are constituents of thought contents, or constituents of abilities to entertain, or reason with, contents. Your belief that this is an encyclopedia entry has the content: *that this is an encyclopedia entry* and that content embeds constituent concepts: *encyclopedia* and *entry*. Your ability to entertain or reason with the content *that this is an encyclopedia entry* involves abilities to think about encyclopedias and entries and, thus, possession of concepts of encyclopedias and entries. This entry restricts attention to a brief discussion of three large questions about connections between concepts and language; it focuses on philosophical rather than psychological issues.

Q1: We use language to express thoughts, and thus, concepts. How are the properties of language that underwrite expression connected with the thoughts, and thus concepts, expressed?

Q2: Humans are unusual, perhaps unique, in both linguistic and conceptual abilities. To what extent are these abilities interdependent? Could a creature think at all—possess any conceptual abilities—without language? Could creatures without language use the same concepts as the linguistically competent? Could a creature have language without concepts?

Q3: Human languages, and human linguistic abilities, appear to vary, for instance in the forms of classification imposed by their vocabularies. Does such variance induce, or reflect, variance in conceptual abilities?

**Linguistic Expression of Concepts (Q1)**

Many philosophers hold the following: (a) Thoughts (e.g., beliefs), in concert with the way the world is, determine truth values, so are either true or false. (b) Determination of truth value for thoughts is derivative from determination of truth value for thought contents (so that, for example, your belief that this is an encyclopedia entry is true because it is true *that this is an encyclopedia entry*). (c) Determination of truth value for thought contents is derivative from contributions made by constituent concepts and the way those concepts are combined in constituting the thought content. From (a) through (c), if we hold fixed the way the world is, then two beliefs with contents constituted in the same way from the same concepts must have the same truth value.

The most straightforward view (SV) about the way language expresses concepts is this: Each substantive expression type (perhaps as typed by meaning) expresses a concept, and each type of sentential mode of combination (broadly, each syntactic structure type, including nonsubstantive expression types) expresses a way of putting concepts together to constitute a thought content. (Very roughly, substantive expression types—such as *heavy* and *encyclopedia*—make a distinctive contribution to the subject matter of a sentence, while nonsubstantives and structural features of sentences—perhaps including *is*, *the*, and the way expressions are combined in “The encyclopedia is heavy”—determine...
how the substantives work together to determine the subject matter of sentences.)

Given a through c, we can test SV by seeing whether, if we hold fixed the way the rest of the world is, all thoughts expressed by a sentence type are guaranteed to have the same truth value. If they are not, then we must reject SV. (If we do not hold fixed the way the rest of the world is, changes in truth-value are to be expected: The true thought that the encyclopedia is heavy would be false if the world was such that it was a far lighter book.)

An immediate problem for SV is the existence of context sensitive expression types, such as indexicals (e.g., I, here, now) and demonstratives (e.g., this, that, those, she, he).

1. I live near London.

Sentence 1 as used by different speakers can determine different truth values. Given a through c, it follows that uses of Sentence 1 can express different thought contents, so SV is false.

A modest revision of SV would restrict it to context-insensitive expression types. However, it is not clear that there are such expression types. Consider Sentence 2:

2. Barack Obama is 1.87 meters tall.

Even if the expression type “Barack Obama” is used only to speak about the current president of the United States (from noon, 20th January 2009), it’s plausible that Sentence 2 can be used to express thought contents with different truth values. For most ordinary purposes, Sentence 2 expresses a truth even if Obama is 1.8700001 meters tall. But plausibly, Sentence 2 would express a falsehood if the context was a precise (nine decimal places) laser measurement of Obama’s height.

A different reason for rejecting SV arises from arguments that concepts are typed in a more fine-grained way than expression meanings. An example of Saul Kripke’s is illustrative. Pierre, a monolingual French speaker, assents to Sentence 3 on the basis of reading travel guides:

3. Londres est jolie.

Pierre moves to an ugly part of London, without realizing it’s the place his travel guides described. Having acquired English from locals, Pierre dissents from Sentence 4:

4. London is pretty.

It is plausible that the thought content Pierre expresses through assent to Sentence 3 is different from the thought content he expresses through dissent from Sentence 4. For otherwise, Pierre would appear to believe and disbelieve the same thought content, and it is more natural to view Pierre as ignorant rather than irrational. The difference in thought contents plausibly traces to a difference in the concepts that Pierre expresses using Londres and London. Nonetheless, it is plausible that London and Londres have the same linguistic meaning. Hence, we have reason to reject SV.

Linguistic Ability and Conceptual Ability (Q2)

Two very general questions about the connection between linguistic ability and conceptual ability are the following. First, is linguistic ability necessary for conceptual ability, so that one could not have conceptual ability without having linguistic ability? Second, is conceptual ability necessary for linguistic ability? Answers to either question depend on operative conceptions of linguistic and conceptual ability.

Many philosophers hold that linguistic ability is bound up with conceptual ability, though Donald Davidson and Michael Dummett are the most prominent defenders of the view. Such views vary in plausibility with variance in the operative conception of language. For example, natural languages—the languages that humans acquire facility with through normal exposure by the age of 3—are governed by rules of case. In English, these rules permit Sentence 5 and rule out Sentence 6:

5. He showed her the encyclopedia.
6. Him showed she the encyclopedia.

The operation of such rules plays no role in arguments for the dependence of conceptual abilities on linguistic abilities; moreover, it is implausible that a creature that lacked sensitivity to case would be precluded thereby from possession of conceptual abilities. Nonetheless, insensitivity to case would evince lack of linguistic ability, as
ordinarily understood. Those who seek to argue that conceptual ability is dependent on linguistic ability should therefore carefully specify the operative features of linguistic ability.

Davidson’s central argument has three main premises. First, having beliefs—or more generally, states with content—requires conceptual grasp of the distinction between beliefs being true and their being false. Second, conceptual grasp of the distinction between truth and falsehood for beliefs requires possession of a concept of belief. Third, grasp of a concept of belief requires linguistic ability. Davidson’s argument for the third premise is not transparent. His idea appears to be that intelligible application of a concept of belief requires a means of grounding the attribution of false beliefs. Unless there is need to appeal to false beliefs in characterizing a subject’s mental life, appeal to belief would be redundant. One might instead appeal to the way the world is. And, the only means of grounding the attribution of false beliefs is to appeal to a feature of the subject’s behavior that is assessable for truth or falsehood in partial independence from the way the world is. Since Davidson holds that only language provides the kind of stable, repeatable feature of subject’s behavior that can be assessed in the required way, he holds that grasp of a concept of belief requires linguistic ability.

Davidson’s argument can be challenged at almost every step. In particular, it is not obvious that having beliefs requires a conceptual grasp of the distinction between true and false beliefs. It is plausible, for example, that infants have numerous beliefs before they are in a position to recognize that some of those beliefs are false. And it is not obvious that appropriate attribution of false belief requires linguistic ability. It is plausible, for example, that prelinguistic infants’ reactions to apparently unexpected outcomes—for instance, a moving ball that fails to emerge in a natural way from behind a screen—can provide nonlinguistic evidence for their having false beliefs.

Davidson’s argument attempts to derive a global conclusion about the connection between conceptual ability and linguistic ability from a particular case, the concept of belief. Increasingly, theorists have shied away from the global question and focused instead on whether linguistic ability is required for possession of particular concepts. Here, we should distinguish between two questions: (1) Is linguistic ability or ability with a particular range of expressions, a requirement on anyone’s ability with a particular concept (e.g., of belief), encyclopedias, Tuesdays, elm trees, and so on? (2) Could linguistic ability be a requirement on someone’s ability with a particular concept? One case where an affirmative answer to (1) is plausible is concepts of days of the week. Plausibly, one could not have a concept of Tuesdays without having a system of seven symbols correlated with day-length periods of time. With respect to (2), one case for which an affirmative answer is plausible is a concept of elm trees. Plausibly, someone might be very bad at telling elms trees apart from oak trees and yet still have distinct concepts of elms and oaks because they stand in the right sort of linguistically mediated relations with experts at distinguishing elms from oaks (a development of a case from Hilary Putnam).

With respect to the second question, René Descartes argued that the exercise of some forms of linguistic ability furnished decisive evidence that the possessor of that ability had a mind and, thus, conceptual abilities. The form of linguistic ability about which Descartes made this claim was the ability to use language appropriately without that use being a merely triggered response to environmental contingencies (e.g., a thermostat’s response to changes in ambient temperature). Alan Turing held a similar view. According to Turing’s Test, ability to engage appropriately in ordinary conversation is a demonstration of mindedness. The plausibility of such views depends on the operative conception of linguistic ability, in particular that it embeds an appropriateness condition. Similar views based on less demanding conceptions of linguistic ability—such as the ability to ape others’ utterances (something a mere recording device might be able to do) or to respond to presentations of colored tiles with utterance of the right color word (an ability akin to that of a complex thermostat)—are less plausible. Plausibly, something close to conceptual ability is built into the operative conception of linguistic ability.

Linguistic Determinism (Q3)

Linguistic determinism is the view that conceptual abilities are shaped by linguistic abilities, so that differences in language can make for differences in conceptualization. Historically, the view is most often associated with Benjamin Whorf’s now largely discredited claim that the Hopi, an indigenous
people of America, spoke and so thought about time in a radically different way from Europeans. The view comes in a variety of strengths, with the strongest forms involving incommensurability between the conceptual repertoires of speakers of different languages so that there is no thought content shared among speakers of different languages, so no agreement or disagreement between them. The strongest form of the view has not found favor in recent times, in part because the strongest form of view cannot be stated or conceived from within the confines of a single linguistic community and (so) is not clearly intelligible. However, research into the role of linguistic ability in local shaping of particular ways of thinking—for example, thinking about color, space, time, and motion—forms an important strand of current research among philosophers and psychologists.

Guy Longworth

See also Classical Theory of Concepts; Concepts, Comparative Perspectives; Concepts, Development of; Concepts, Philosophical Issues

Further Readings


Types of Combinations

Cognitive psychologists and linguists have studied how people interpret noun-noun combinations. Although there are a great many ways that people interpret noun-noun combinations, there is consensus among researchers that there are three general types of interpretations: relation, property, and hybrid. A relation interpretation involves a relation between two nouns. For example, a person might interpret robin hawk as a “hawk that preys on robins.” In this case, the robin and hawk play different roles in the relation. The hawk plays the role of a predator and the robin plays the role of the prey. A property interpretation involves attributing a property of the modifier to the head noun. For example, a “frog squirrel” could be interpreted as “a squirrel that hops” (inheriting the property of hopping from frog). A hybrid interpretation refers to a mixture of the modifier and the head noun, as in “cow horse” or a cross between cow and a horse.

Interpreting Combinations

Three types of combinations have been identified. But how does a person determine a plausible meaning? One strategy is to interpret the combination based on context. For example, an author of a magazine might refer to a “scorpion fly” and describe its meaning as a “fly that has a tail similar to a scorpion.” In this case, the author is describing
a property interpretation. A person familiar with a scorpion then can easily interpret the meaning of scorpion fly. However, unpublished research in Ed Wisniewski’s laboratory suggests that often there is not much explicit context provided for the reader. Instead, a person often must determine the meaning or the interpretation of the combination from the two nouns. According to Greg Murphy, people, in many cases, rely on their background knowledge to interpret meanings of combinations. For example, a reporter for a newspaper might refer to a novel combination such as “lima bean casserole” without describing its meaning. However, even if a person may have never heard of lima bean casserole, most people are likely to interpret the combination as “a casserole made of lima beans” (a relation interpretation). For example, most people know that lima beans are food, that casseroles are baked, and so on. They use their background knowledge of the words to determine the meaning of the combination.

There are other factors, in addition to background knowledge or context, that affect how a person interprets a noun-noun combination. When two nouns of a combination are very similar, people are much more likely to give a property interpretation to the combination (e.g., “whiskey wine” interpreted as “a stronger type of wine”). One reason for the preponderance of property interpretations for highly similar combinations is that the nouns of relation interpretations have to play different roles. If two nouns are very similar, then typically they cannot play different roles. In the combination whiskey wine, both nouns are so similar that they play the same role (“to drink”), preventing a relation interpretation and leading to the property interpretation.

Models of Conceptual Combination

A number of models of conceptual combination have been developed by researchers and will be briefly described.

Christina Gagne and Edward Shoben have developed a model of a conceptual combination that focuses on relation interpretations. Their model shows that a combination is easier to understand based on the frequency of the relation interpretations in language.

Zachary Estes and Sam Glucksberg’s interactive property attribution model applies to property interpretations. In their model, property interpretations occur when a salient modifier is activated and attributed to the head noun. For example, in the combination “zebra bag,” a salient feature of zebra is “having stripes.” This feature is then attributed to the head noun bag.

James Hampton’s attribute inheritance model applies to hybrids. It emphasizes the emergent features that arise from the two nouns when they are combined. For example, Hampton found that in “pet bird,” pets that are also birds (hybrid) were judged to have the emergent features “live in cages and talk” that are false for pets or birds considered separately.

All of these models address only one of the three different interpretations. However, Wisniewski’s dual-process model can account for relation, property, and hybrid interpretations. One process produces relation interpretations; another process produces property interpretations and hybrids.

Conceptual Combination in Another Culture: Chinese

Studies have demonstrated that noun-noun combination is also a very productive way for generating new words in the Chinese language. Jing Wu conducted a survey of the Dictionary of Contemporary Chinese New Words. She found that among 15,400 new words and expressions, 25% of the new phrases are noun-noun combinations.

Wu’s study showed that the three major types of noun-noun combinations in American English also characterize those in the Chinese language. Relation interpretations are most common, followed by property interpretations. Hybrid combinations are a rare type in Chinese, too. For example, the ideograms meaning lion tiger refer to a cross between a lion and a tiger. Relation interpretations included color TV wall, meaning “a giant thin screen of color TV installed in a wall”), and mother’s milk bank, a bank for storing mothers’ milk. Property interpretations included bubble economy, an economy that does not last as long as a bubble, and toad glasses, eyeglasses with large frame and glass that look like the eyes of a toad.

Emergent Features in Conceptual Combination

In general, emergent features are not initially present in concepts but arise when these concepts are
Conduction Aphasia is an acquired disorder of language typically caused by stroke. Classically, the symptoms of conduction aphasia—impaired verbatim repetition, phonemic errors in speech production, and naming difficulty—have been attributed to a disconnection between speech reception and production areas. However, recent work suggests that conduction aphasia is an auditory-motor integration disorder caused by damage to a cortical network in the left posterior Sylvian region. This entry reviews the symptoms and neurology of conduction aphasia, the classical interpretation of the disorder, and more recent functional-anatomic accounts.

Characterization of Conduction Aphasia

Conduction aphasia, first identified in 1874 by Carl Wernicke, is an acquired disorder of language, typically caused by stroke, that is characterized by frequent phonemic paraphasias (sound-based speech errors) with attempts at self-correction, impaired verbatim repetition, naming difficulties, but otherwise fluent and grammatical speech output. Auditory comprehension is relatively spared. The deficits in conduction aphasia are exacerbated with increasing phonological load such as when the task involves multisyllabic words, phrases, and/or reduced semantic content. The three major symptoms all point to a deficit at the phonemic level of processing. The paraphasias tend to be phonemic level errors...
(sound based errors such as \textit{strool} for \textit{stool} or \textit{okie} for \textit{cookie}), the repetition deficit appears to reflect dysfunction of phonological short-term memory (ability to immediately recall a list of words, numbers, or nonsense words), and the naming deficit is both aided by phonemic cueing (providing the first sound) and characterized by tip-of-the-tongue states, that sense of knowing the word or name one is looking for but failing to access it, which has been linked to failures to access phonological information during naming. Thus, while conduction aphasia is commonly referred to as a disorder of “repetition,” it is clear that the syndrome is not restricted to this one behavior. Instead, the disorder appears to involve a phonemic level of processing that is appreciable not only in repetition but also in other production-related tasks. Crucially, the phonemic level deficit does not substantially affect receptive functions as auditory comprehension is typically well preserved in conduction aphasia. Consistent with this, it has been observed that conduction aphasics often get the gist of sentences that they fail to repeat. For example, one patient, when asked to repeat the phrase, \textit{The pastry cook was elated}, responded with \textit{Something about a happy baker}.

The Classical Interpretation

Classically, conduction aphasia was thought to result from damage to a white matter bundle, the arcuate fasciculus, which resulted in a disconnection of the two major speech centers, Wernicke’s and Broca’s area. This is no longer a viable model, however, because damage to the arcuate fasciculus is not associated with conduction aphasia, and because cortical stimulation of the left posterior temporal lobe is sufficient to cause the symptoms of conduction aphasia. Recent evidence suggests instead that conduction aphasia is caused by damage to the left posterior superior temporal gyrus and/or the left supramarginal gyrus, that is, a location centered around the left planum temporale/parietal operculum (see Figure 1).

Conduction Aphasia as a Short-Term Memory Deficit

Modern theorists have suggested that the repetition deficit in conduction aphasia results from a disruption of phonological short-term memory. On this view, the phonological trace fades too quickly in conduction aphasics to allow for accurate verbatim repetition of speech, whereas comprehension is spared because access to semantic representations is less dependent on phonological working memory. Consistent with this view is the observation that conduction aphasics can be impaired on tests of phonological short-term memory and often get the gist of a sentence without retaining a memory for the specific words used. This clearly explains one aspect of the symptom cluster in conduction aphasia but does not alone explain the other phonological-level symptoms of conduction aphasia such as the naming deficit and phonemic paraphasias. It is possible that the short-term memory deficit and the other phonological symptoms stem from damage to different neural subsystems that happen to be damaged in conduction aphasia.

Conduction Aphasia as an Auditory-Motor Integration Deficit

An alternative possibility that incorporates the phonological short-term memory hypothesis but also explains some of the other symptoms of...
Confabulation

One of the Latin roots of the term _confabulation_ is _fabulari_, which became our word _fable_. When the German neurologists Karl Bonhoeffer, Arnold Pick, and Carl Wernicke began using the term in the early 1900s, they applied it to false memory reports, little fables, created by their patients, who suffered from a syndrome that later came to known as Korsakoff’s amnesia. When asked what they did yesterday, these patients do not remember, but will report events that either did not happen or happened long ago. The technical definition of confabulation the early neurologists coined has three components: (a) Confabulations are false, (b) confabulations are reports, and (c) confabulations are about memories.

During the remainder of the 20th century, however, the use of the term was gradually expanded to cover claims made by other types of patients, many of whom had no obvious memory problems, including patients who deny illness, split-brain patients (who have had their hemisphere surgically separated to prevent the spread of epileptic seizures), patients with misidentification disorders (who make false claims about the identities of people), and patients with schizophrenia, as well as children and normal adults in certain situations. Patients who deny that they are paralyzed have been claimed to confabulate when they provide reasons for why they cannot move (“My arthritis is bothering me,” “I’m tired


**Confabulation**

Conduction aphasia has been suggested by recent studies investigating a sensory-motor integration network for speech and related functions. Sensory-motor integration has been studied extensively in the visual-motor domain, and a set of brain regions have been identified in the posterior parietal lobe of both monkeys and humans that appear to perform a transform function mapping between visual and motor representations. These areas show both sensory and motor response properties, are densely connected with frontal-motor areas, and are organized around particular motor effector systems (e.g., eyes vs. hands). Recent work in humans has identified a similar sensory-motor integration network for speech functions (or more accurately, for the vocal tract motor system). Functional imaging has found that a left dominant region within the Sylvian fissure at the parietal-temporal boundary (area Spt) exhibits the same response properties found in sensory-motor integration areas in the posterior parietal lobe. It has been proposed that area Spt is a central hub in a sensory-motor integration network that transforms auditory-based representations into vocal motor-based representations.

Interestingly, the location of area Spt at the parietal-temporal boundary is in the center of the lesion distribution associated with conduction aphasia. Furthermore, Spt activates during the performance of behaviors that are impaired in conduction aphasia (repetition, naming, short-term memory). For these reasons, and because Spt’s functional properties show parallels to other known sensory-motor integration areas in the posterior parietal lobe, it has been suggested that conduction aphasia is a deficit caused by damage to area Spt and surrounding tissue, which disrupts the interaction between sensory and motor systems involved in speech.

_Gregory Hickok_

_Author’s Note:_ This work was supported by NIH grants R01 DC03681 & R01 DC009659.

_See also_ Aphasia; Production of Language; Working Memory in Language Processing

_Further Readings_

of following your commands”). Another type of patient will deny blindness and attempt to answer questions about what he or she sees, producing what have been called confabulations (“It’s too dark in here”). Misidentification patients have been said to confabulate when asked what the motives of the “impostor” are, or why someone would go through all the trouble to impersonate someone else (“Perhaps my father paid him to take care of me”). Similarly, when the left hemispheres of split-brain patients attempt to answer questions without the necessary information (which is contained in their right hemispheres), this has also been called a confabulation.

There are thus currently two schools of thought on the proper scope of the concept of confabulation: (a) those who remain true to the original sense and so believe that the term should only be applied to false memory reports and (b) those who believe that the term can be usefully applied to a broader range of disorders. This possible expansion of the concept forces several difficult questions. Has the concept expanded so much as to become meaningless? Do the new confabulation syndromes share anything significant with the classical memory cases? Minimally, some confabulation syndromes involve memory (Korsakoff’s, aneurysm of the anterior communicating artery), whereas others involve perception (denial of paralysis or blindness, split-brain syndrome, misidentification disorders). This entry focuses on the broader concept of confabulation and includes a description of confabulations based on both memory and perception. It will also examine what is known about the neuropsychology of confabulation.

The Broader Sense of Confabulation

The following definition is based on the idea that confabulation syndromes involve malfunctions in different knowledge domains, as well as executive system damage.

Jan confabulates that $p$ (where $p$ is some proposition) if, and only if

1. Jan claims that $p$.
2. Jan believes that $p$.
3. Jan’s thought that $p$ is ill grounded.
4. Jan does not know that her thought is ill grounded.
5. Jan should know that her thought is ill grounded.
6. Jan is confident that $p$.

“Claiming” covers a wide variety of responses by subjects, including drawing and pointing. The second criterion expresses the sincerity of confabulators. The third criterion refers to the problem that caused the flawed response to be generated; this problem is the first factor. The fourth criterion refers to the failure of the second phase, the failure to reject the flawed response (the second factor). The fifth criterion captures the normative element of our concept of confabulation. If the confabulator’s brain were functioning properly, he or she would not make that claim. The last criterion refers to another phenomenon typically seen in confabulators, the serene certainty they have in their communications, which may be connected to the frequent finding of low or abolished sympathetic autonomic activity in confabulating patients.

Types of Confabulation

Confabulations About Memories

The presence of confabulation is a defining characteristic of two memory syndromes, Korsakoff’s syndrome and a similar syndrome caused by aneurysm of the anterior communicating artery, an artery that forms the anterior portion of the circle of Willis. Alzheimer’s patients will also frequently confabulate about memories, and young children are also prone to reporting false memories. These four types of confabulators (Korsakoff’s patients, anterior communicating artery patients, Alzheimer’s patients, and young children) have an initial memory retrieval problem, coupled with a failure to check and correct their false memories. Apparently the children’s prefrontal areas have not yet fully developed, while the Alzheimer’s patients’ prefrontal lobes have been damaged by the amyloid plaque lesions. In contrast, there exist many memory patients with damage restricted to more posterior parts of the memory system (e.g., to the hippocampus or other parts of the temporal lobes) who openly admit that they
cannot remember and are not prone to producing confabulations.

Confabulations About Perceptions

Vision

Anton’s syndrome is a condition in which the patient is partially or entirely blind yet insists that her vision is fine. The posterior damage site typically involves bilateral lesions to the occipital cortex, causing the blindness, coupled with prefrontal damage, apparently causing an inability to become aware of the blindness. Some of these patients appear to be mistaking hallucinations for actual perceptions. Split-brain patients may also confabulate when the left hemisphere is asked about perceptual information contained only in the right hemisphere.

Somatosensation

The patients who deny that they are paralyzed have a condition referred to as anosognosia, meaning unawareness of illness. They typically have damage to one or more lower level somatosensory systems responsible for representing the affected limb, in addition to frontal damage that may compromise areas capable of making the patient aware of the damage to the lower level somatosensory systems.

Person Perception

Perceptual confabulations are also issued by patients suffering from the misidentification syndromes (especially Capgras syndrome, in which the patient claims that people he or she knows well have been replaced by similar looking impostors). Such patients do not perceive people close to them correctly and produce confabulations about impostors as an attempt to explain this. These syndromes also show a pattern of posterior cortical damage, often to the temporoparietal cortex, coupled with prefrontal damage, typically in the right hemisphere.

Confabulations About Intentions and Actions

Patients who have undergone a split-brain operation may confabulate about actions performed by the right hemisphere. In a typical experiment, commands are sent to the right hemisphere only, while the left hemisphere, unaware of this, confabulates a reason for why the left hand (controlled by the right hemisphere) obeyed the command. Similar sorts of confabulations can be elicited by brain stimulation. For example, the patient’s cortex is stimulated, causing an arm to move. When asked why the arm moved, the patient claims he or she felt like stretching the arm. Hypnotized people may also confabulate about actions. In a typical case, the subject is given a hypnotic suggestion to perform a certain action but then confabulates by offering a different, ad hoc, reason for the action when asked.

Confabulations About Emotions

False or ill-grounded self-attributions of emotions are another type of confabulation. For example, in one experiment, people were given an injection of adrenaline without their knowledge but attributed their inability to sleep to things such as uneasiness about a romantic relationship or nervousness about what they had to do the next day.

The Neuropsychology of Confabulation

There are several clues as to the nature and location of the neurological damage in confabulation patients: (a) The patients with aneurysms of the anterior communicating artery—a tiny artery near the anterior commissure that completes the anterior portion of the Circle of Willis—provide our best clue about the locus of the frontal problems in memory confabulation. This artery supplies blood to portions of the posterior orbitomedial cortex known to have mnemonic functions. (b) Split-brain patients confabulate about information perceived by the right hemisphere. The right hemisphere, or lack of communication with the right hemisphere, shows up in several of the perceptual confabulations. This may indicate that confabulations are created by the left hemisphere in the absence of disconfirming information possessed by the right hemisphere.

William Hirstein

See also Amnesia; Anosognosia; Capgras Delusion; Delusions; Fregoli Delusion

Further Readings


**Conscious Thinking**

This entry addresses the question of the nature of thought and its relation to consciousness. Current philosophical orthodoxy holds that thought and consciousness are only contingently related. Though thoughts *may be* conscious, it is not in their *nature* to be such. The problem of thought and the problem of consciousness—for most philosophers, how they fit into a naturalistic worldview—are thus wholly distinct. Hence, the present intractability of the latter problem does not constitute an in-principle barrier to a scientific solution to the former. Reasons are reviewed here for thinking that, on the contrary, thought and consciousness are inextricably linked and, thus, that the prospects for naturalistic explanation of thought are not as good as some have thought.

**Intentionality and Rationality**

Late 20th-century theories of thought and thinking in the analytic philosophical tradition focused on the problem of how it is that presumably purely physical beings such as ourselves can be in states (or have brains that are in states) that are *about* things or have “intentionality” and that can bear logical relations to each other such that sequences of them may be *rational* or *irrational*.

To say that a mental state is *intentional* is to say that it has a *content*—something analogous to the *meaning* of a sentence. The term *intentional* is used here in application only to so-called propositional attitudes (beliefs, desires, hopes, etc.) and the constituent thoughts that render them intentional (what they are about), skirting the issue of the intentionality of perceptual and other sensory experiences. The sentence “Blood is red,” for example, has (in English) a particular meaning and in virtue of this meaning (derived from the meanings of its constituent terms) is *about* blood and says *that* it is red. The sentence “Mud is brown,” in contrast, has a different meaning and is about something else and says something different about it. Further, these sentences have, in virtue of their meanings, *truth conditions*—that is, they specify the worldly conditions under which they are true (or false)—as well as *logical properties*—relations of consistency, inconsistency, and entailment to other sentences. The sentence “Blood is red” is true if and only if blood is red; otherwise, it is false; “Blood is red” and “Mud is brown” are logically consistent with each other (they can both be true); “Blood is red” and “Blood is brown” are not consistent with each other (they can’t both be true); and “Blood is red and mud is brown” logically entails “Mud is brown.” Exactly analogous things may be said about the thoughts that blood is red and that mud is brown. Indeed, it is a traditional assumption in analytic philosophy of language that the meanings of sentences derive from (or are) the contents of the thoughts they are (by convention) used to express. Thus, thoughts have contents, which determine their truth conditions and logical relations to each other. (Jerry Fodor has championed the view that we think in a “language of thought,” a system of symbolic representations with sentence-like structures.)

**Naturalism**

Philosophical theories of intentionality and rationality have typically been committed to “naturalism” (the view that these phenomena can be explained in terms consistent with the natural sciences—in
particular, neurophysiology, biology, and ultimately, physics) and have typically sidestepped the question of consciousness. Attempts to explain the propositional attitudes (belief, hope, desire, fear), their contents (what they are beliefs in, hopes or desires for, fears of), as well as the logical relations among them, have generally not taken account of the fact that some of them are conscious. The reason for this is twofold. On the one hand, since there’s very good evidence that there are unconscious thoughts and thought processes, it would seem that what makes a mental state a thought has little, or even nothing, to do with what makes it conscious. On the other hand, this is a lucky break, since no one has the slightest idea what consciousness is or how to provide an explanation of how it could arise from brain activity. If intentionality and consciousness can vary independently of one another, the latter may safely be ignored when theorizing about the former. Whatever explanation there may be for consciousness generally may simply be combined with whatever theories of intentionality and rationality turn out to be the right ones in order to explain conscious thought and thinking.

In addition, there has been widespread optimism about the feasibility of naturalistic explanations of rationality and intentionality.

Rationality

The conception of the mind as a kind of computer, and of thinking as a kind of rule-governed symbol manipulation, gained new life in philosophy through the work of Hilary Putnam, who hypothesized that the mind is a kind of Turing machine. Alan Turing showed how a merely material device—and one of no particularly special or interesting physical type—could engage in characteristically intellectual operations such as addition, subtraction, multiplication, and division. By simply reading, writing, and replacing symbols, such a machine could convert a representation of any mathematical problem into a representation of its solution. (The construal of thinking as symbol manipulation was also endorsed by Thomas Hobbes.) A thing need not have an immaterial soul in order to have a mind. Turing himself argued that a computational device that was conversationally indistinguishable from a human being would thereby literally deserve the honorific “thinker.” Contemporary cognitive science has explicitly extended the realm of the computational (at least in principle) to virtually every “mental” operation.

Intentionality

The problem of intentionality—the problem of how it could be that a physical object could be in states that are about things—is also widely believed to have been (as far as philosophers are concerned) essentially solved by an approach introduced into the philosophical literature by Fred Dretske. Dretske showed how a merely material object could be in states that carry information about the existence or condition of other things and argued that the property of carrying information is a kind of proto-intentionality. The rings of a tree, for instance, represent the tree’s age; the presence of smoke indicates the presence of some sort of combustion; the occurrence of a certain sort of spots on the skin means measles. These relations, which hold in virtue of lawful relations of cause and effect between the phenomena, are sufficiently like the “aboutness” of intentionality (witness the language we use to describe them) to suggest that they could be the basic materials from which genuinely intentional systems such as ourselves might be constructed (by, e.g., evolution). The basic ingredients of intentionality are ubiquitous in the natural world.

Of course, trees, clouds of smoke, and skin don’t actually think; so there’s a good deal more work to be done to explain how these basic ingredients are exploited to produce genuine minds. Characteristically, philosophers disagree, vigorously, about how—and whether—this is to be done. Quite different theories based on Dretske’s basic insight have been proposed by, among others, Ruth Millikan, Daniel Dennett, Fodor, and Dretske himself. The central problem has been to distinguish intentional systems such as ourselves, which are capable of states that misrepresent the world, from purely informational systems, which cannot. Since an effect cannot occur uncaused and is, necessarily, caused by whatever caused it, it is not possible for it to carry misinformation about its cause. To do so, it would have to be caused by something that did not cause it.

These conceptual breakthroughs have been a source of excitement and genuine hope that the longstanding mystery of what our minds (our cognitive minds, at least) are and how it is that we have them might be solved within the bounds of natural
Conscious Thinking

Our phenomenal minds—our conscious experiences, with their qualitative characters—may remain a mystery (but see below); yet given that the cognitive and the conscious are metaphysically distinct, we may nonetheless hope for a scientific explanation of a significant portion of our mental lives.

Consciousness

There are those, however, who deny that the problem of cognition (intentionality and rationality) and the problem of consciousness (qualitative experience) can be segregated in this way. They persist in the Cartesian intuition that, somehow, there cannot be thought—or cannot really be thought—or thinking in the absence of consciousness. (René Descartes himself held that consciousness and mentality are coextensive; that is, all and only conscious states and processes are mental.) John Searle, with his connection principle is probably the best known of these, though there are others, such as Galen Strawson and Charles Siewert. For such philosophers, there is something about what makes a state conscious that's essential to making it intentional.

Phenomenology of Thinking

One way of substantiating the Cartesian intuition is to provide reasons for thinking that there is a "phenomenology" of conscious thinking, which is essential to its identity as thinking. To say that a state or process has a phenomenology is to say that it has features in virtue of which there is "something it's like" (in Thomas Nagel's phrase) to be in it or to undergo it. For example, the experience of being in pain has a distinctive sort of qualitative character (phenomenology), which is quite different from that of the experience of, say, hearing thunder or tasting chocolate: What it's like to be in pain is quite different from what it's like to hear thunder or to taste chocolate. The thesis that there's an essential phenomenology of conscious thought holds that there is something it's like to consciously think a thought, which is distinctively cognitive (i.e., not visual or auditory or olfactory or gustatory or . . . ) and which individuates the thought (makes it the thought that it is and distinguishes it from other thoughts), in a way analogous to that in which, say, the distinctive auditory phenomenology of the sound of thunder distinguishes it from the sound of sleigh bells, or the distinctive gustatory phenomenology of the taste of chocolate distinguishes it from the taste of garlic. What makes a conscious mental state a thought is its having a distinctive phenomenology of the cognitive sort.

Unconscious Thoughts

Given the widely shared (but not inevitable) intuition that in order for a state to have phenomenality (i.e., to be a phenomenal state) it must be conscious, the view that there is an essential phenomenology to thinking seems to have the problematic consequence that there can be no unconscious thoughts or thinking. Given what we have learned from Sigmund Freud and from contemporary cognitive science about the existence and, perhaps, primacy of unconscious mentation, however, it seems intellectually recidivist (at least) to advocate a return to a Cartesian view of the mind. Searle and Strawson have attempted to tackle this problem—Searle with the view that unconscious intentional states are, necessarily, potentially conscious, and Strawson with the view that though a particular state need not be conscious in order to be intentional, an unconscious state can only be intentional if it's a state of a creature capable of being in conscious states. Neither of these views seems completely satisfactory. (What is the property that unconscious states have that can render them intentional? Why should the intentionality of an unconscious state depend on the consciousness of some other state?) Though there are other ways one might try to face down the problem. For example, one might simply bite the bullet and deny that in general consciousness is necessary for phenomenality. This would of course break the connection between thought and consciousness, but not between thought and something very close to consciousness—namely, phenomenality, a necessary condition for it. (One might also deny that states conscious in themselves need be conscious for their possessor.)

Representationalism

One popular approach to the phenomenology of conscious states—representationalism (or intentionality)—might be thought to be of some use here. Representationalists such as Dretske, Michael Tye, and William Lycan hold that the phenomenal contents of conscious mental states are a species of intentional contents. On this view, the qualitative features associated with a conscious perceptual state,
those one would mention in characterizing what it is like to be in that state, are the qualitative features of the thing(s) perceived. They come to be associated with the perceptual state in virtue of the latter’s representing them, in a manner not essentially different from that in which intentional states such as thoughts and beliefs represent their contents. The styles of representation may be different (e.g., perceptual representations may have an image-like structure, while cognitive representations may have a sentence-like structure) but the features that determine which properties are represented are of the same basic type in both cases, namely, “tracking” or informational relations, of essentially the type suggested by Dretske. The qualitative feature one might mention in characterizing one’s visual experience of the sky at noon on a clear day, for example—the blueness—is a property of the sky, not of one’s experience. (To maintain otherwise is to commit what U.T. Place termed the “phenomenological fallacy.”) The only qualitative properties there are, are the qualitative properties of extramental objects; so there is no special problem of explaining how a mental state could “have” conscious qualitative character.

Moreover, the explanation of the qualitative character of experiences on this type of theory is independent of the explanation of consciousness. Typically, consciousness is explained in terms of first-order representation—a creature is conscious of a thing x if and only if it is in a state that represents x—where this first-order representational state need not itself be conscious (in the sense that its possessor need not be aware of being in it; such awareness requires a higher order representation). Thus, on the representationalist account qualitative character and consciousness are metaphysically independent.

Representationalist approaches face two serious problems, however, one internal and the other in application to the present problem. The internal problem is the explanation of the qualitative character of dreams and hallucinations—experiences in the absence of instantiated properties to represent. Though this problem has been addressed (by, among others, Dretske, Tye, and Lycan), it’s not clear that it’s been solved. In the context of phenomenally constituted thought content, the other problem is the identification of objective qualitative properties to serve as cognitive phenomenal character.

David Pitt

See also Access Consciousness; Causal Theories of Intentionality; Consciousness and Embodiment; Consciousness and the Unconscious; Disjunctive Theory of Perception; Representational Theory of Mind; Thinking

Further Readings

CONSCIOUSNESS, COMPARATIVE PERSPECTIVES

Do animals have mental experiences or are they closer to robotic zombies? Most animal behaviorists these days, at least those steeped in understanding evolutionary continuities in brain and mind organization, are happy with the provisional conclusion that many other animals, surely mammals and birds, perhaps even some invertebrates (octopuses), are conscious beings. However, few would be able to outline how one could credibly (scientifically) study such aspects of animal minds. The perennial problem, as noted by ethologist Niko Tinbergen (1951) in his seminal treatise, The Study of Instinct, was simply “Because subjective phenomena cannot be observed objectively in animals, it is idle to claim or deny their existence” (p. 4). With the advent of modern neuroscience, that statement no longer has the ring of truth that it used to, even though a series
of famous animal behaviorists still tend to subscribe to Tinbergen’s dictum.

Of course, the study of experience in animals had to wait for the maturation of neuroscience in the psychological and animal behavioral sciences, which started in earnest in the early 1970s. Before then, compelling experimental strategies could not be generated to study the nature of animal experiences at a scientific-causal level, even though there had been abundant treatises considering such issues at conceptual-inferential levels (not considered here). We can now be confident that all mammals and birds are beings that have internal experiences, especially relevant for engendering the qualities of “rewards” and “punishments” they encounter in the world (i.e., affective consciousness, which can be monitored by the rewarding and punishing states evoked by artificial activation of brain emotions stimulation with direct electrical or chemical stimulation of the brain). The existence of their perceptual experiences (i.e., cognitive consciousness) is likely but harder to understand neuroscientifically, since there are no comparable evaluative measures and one has to infer experience simply from an animal’s capacity to discriminate stimuli. This tells us little about the experience itself. Thus, this entry will focus on affective consciousness—the experienced feeling of emotions—rather than perceptual or cognitive consciousness—the felt experience of the world, since the neuroscientific evidence is much stronger for the former than the latter. However, Don Griffin and Bjorn Merker have made a strong and reasonable case for the conclusion that animals do have subjective cognitive experiences of the world.

Why is a resolution of the opening question so important? Perhaps most poignantly, it could modify how we envision animals as fellow creatures, leading hopefully to more sensitive practices in animal research, animal husbandry, and our reverence for life. Beside such practical and philosophical issues, a scientific understanding of “lower order” (more primal) core forms of consciousness in animals is culturally important and helps set the stage for deeper understanding of human consciousness than is currently possible through research on people, where the detailed neuroscientific work simply cannot be done. Thus, a scientifically satisfactory understanding of higher order consciousness (“awareness”) may not be achieved without illuminating the foundations of subjective experience in the animal kingdom.

Ancestral Sources of Mind—From “Experience” to “Awareness of Experience”

There are good reasons to suspect that “affective consciousness” was among the earliest experiences in BrainMind evolution—a term that will be used here interchangeably with MindBrain to highlight a thoroughly monistic view of mental life. Monism acknowledges that mental (mind) experience is bound to brain functions—they are two aspects of the same physical processes—as opposed to dualism, which sees the brain and mind as thoroughly separate. Abundant evidence now indicates that raw emotional experiences arise from ancient (subcortical) brain networks, shared homologously by all mammals and birds, that control primary-process emotional behaviors—namely, brain processes that generate what used to be called “instinctual” emotional behaviors by ethologists, and “unconditioned stimuli and responses” by behaviorists. This robust fact led to dual-aspect monism research strategies, ones that recognize that instinctual emotion behaviors are evidence of affective experience. For example, a cat that has its ears laid back and is hissing, is not just producing an innate behavior with no corresponding feeling, it is actually experiencing RAGE (we would say “anger” but the use of vernacular terms has many problems in cross-species neuroscience). Thus, despite Tinbergen’s caution, spontaneous (unlearned) emotional behaviors can currently be used as validated proxies for emotional feelings in animals.

Another key point, as one takes an evolutionary perspective to basic emotions, is that the brain is the only organ of the body where evolutionary passages (layers of control) remain evident at neuroanatomical and functional levels. Accordingly, Jaak Panksepp reasons that levels of human consciousness may be fathomed by understanding the successive waves of brain evolution that provided for the increasing complexity of animal minds. He argues that many contentious and paradoxical issues in behavioral and psychological sciences—often caused by the inability to generate adequate agreed-upon definitions for key concepts—diminish as we seriously envision the evolutionary layering of the BrainMind: A multitiered, cross-species BrainMind.
approach to consciousness can now envision a tripartite gradation of MindBrain evolution.

Panksepp further posits that we can evolutionarily divide consciousness (defined as simply having internal experiences) into (a) primary-process core consciousness (e.g., raw brain-bodily feelings controlled by ancient subcortical processes), (b) secondary-process consciousness (e.g., experiences arising from universal learning mechanisms such as Pavlovian/classical conditioning and instrumental/operant learning, and (c) tertiary-process experiences, which recruit higher brain mechanisms, especially neocortex, allowing individuals to reflect, think, and ruminate about their experiences, requiring abundant working memory to retrieve and juggle information for higher order reflections and explicit planning of courses of action.

Endel Tulving, the great cognitive neuropsychologist, envisioned human consciousness in such terms, with (a) anoetic consciousness—experience without understanding; (b) noetic consciousness—knowing facts about yourself and the world; and (c) autonoetic consciousness—being able to use episodic, personal memories of life events to time travel backward and forward mentally, to better understand the past and anticipate the future. These synergistic viewpoints, recognizing “nested hierarchies” in BrainMind organization, as described by Georg Northoff and colleagues, can diminish disputes among those working at different levels of control. Both top-down and bottom-up controls exist in the BrainMind.

How many of these levels of consciousness do other animals have? We do not know. However, Panksepp and colleagues suggest that other animals possess up to all three forms, depending on their level of evolutionary brain development. As they explain, we can have more confidence about the primary-process anoetic forms of consciousness, which may be the evolutionary foundation for the rest. Certainly many animals exhibit sophisticated learning and, hence, probably have secondary-process, noetic consciousness. Perhaps large-brained higher primates and birds (e.g., crows, ravens, parrots) even have tertiary-process autonoetic experiences, but this, they admit, is harder to defend scientifically. However, consider scrub jays that remember who was watching them when they cached food, and recache their treasures when observers are no longer about. In any case, Panksepp and colleagues assert that many animals have affective “experiences”—from pain to joy—without necessarily knowing that they reflect on (are aware of) these experiences. Indeed, most other animals may lack the noetic knowledge—the self-awareness—that they are conscious beings. But this does not make them unconscious.

**Emotional Feelings in Animals**

However, it must still be noted, the question of emotional feelings in animals remains a contested issue. Animal behavior specialists working in a more popular genre, such as Marc Bekoff and Temple Grandin, express no doubts that animals experience and act on feelings. They say we can know their emotional state by their behaviors. Most other researchers now agree that animals have emotional behaviors but hesitate to proclaim that those behaviors are expressions of emotional feelings. In “Who Needs Consciousness?” Marion Dawkins (2001) follows Tinbergen’s lead when she states, “However plausible the assumption that other species have conscious experiences somewhat like ours is, that assumption cannot be tested in the same way that we can test theories about behavior, hormones or brain activity” (p. S28). Similarly, Franz de Waal (2011) encourages scientists to study emotions in ways that “avoid unanswerable questions and to view emotions as mental and bodily states that potentiate behavior appropriate to environmental challenges” since “we cannot know what they feel” (p. 191). All of these scientists base their conclusions on behavior-only evidence, but none have considered the functional characteristics of the emotional circuits of the brain—namely, the rewarding and punishing networks where primal emotional behaviors originate.

The field of neuroscience has opened up opportunities to study emotions in the brains of both human and nonhuman animals, but many neuroscientists still do not agree on whether animals have feelings. Joseph LeDoux has long proposed that feelings emerge from dorsolateral prefrontal cortical working-memory processes, which because of their modest size in most laboratory animals, would presumably mean that animals have no emotional feelings. Similarly, Antonio Damasio has typically claimed that affective feelings are a product of a neocortical read-out process. In his early books, Damasio repeatedly stated that animals have
emotional behaviors but humans have emotional feelings. More recently, he has acknowledged that animals do have emotional feelings though he still concludes that they are a neocortical process. While these conclusions may seem reasonable, they are not based on causal evidence. In addition, they do not explain the evidence that humans and animals with no neocortex are still fully emotional beings. In contrast, Panksepp asserts that both raw emotional behaviors and feelings originate within subcortical systems that are homologous in all mammals. Since all mammals have these homologous systems, all mammals, human and nonhuman, must also experience their emotions, at least as negatively or positively valenced feelings.

Panksepp focuses largely on primary-process anoetic consciousness in animals (pure experience, especially emotional-affective experiences—good and bad feelings), because the neurobehavioral evidence is robust. By using electrical stimulation of the brain (ESB) and measuring behavioral responses, Panksepp determined that there are at least seven subcortical emotional systems: SEEKING, RAGE, FEAR, LUST, CARE, PANIC/GRIEF, and PLAY. The capitalization denotes that these terms are labels for specific brain networks that elaborate specific emotional responses to stimuli. When given the choice, animals turn on ESB to the positive emotional systems and turn off ESB that arouses negative feelings. Similarly, they show conditioned place preference or avoidance to locations where positive or negative arousing ESB was received. Finally, when they receive ESB to these same brain areas, humans report overwhelming emotions appropriate to the system that was stimulated. Thus, according to Panksepp, the evidence indicates that primary-process affective experiences arise from unconditioned emotional response systems of the brain.

Cognitive Processing Without Consciousness?

What may seem unconscious at a high cognitive-conceptual-verbal level (namely, our higher levels of “awareness”) may be experienced at a deep (subcortical) affective level. For example, in their review of “unconscious goal pursuit” research, Ruud Custers and Henk Aarts discuss a study in which participants squeezed a handgrip harder after receiving a subliminal cue associating the squeeze with a larger monetary award. However, participants reported no conscious decision to squeeze harder. In addition, activity increased in brain areas associated with reward processing and action preparation. Using Panksepp’s terminology, these brain changes may represent shifts in primal SEEKING urges. Thus, the participants may have experienced an energized feeling of “wanting” without being conscious of increased desire for anything specific, namely the many objects of desire. Perhaps unreflective (anoetic) experiential changes are often hidden under what is called “conscious awareness.” If so, investigators cannot reach credible conclusions unless they try to evaluate for the existence of anoetic experiential shifts. As we better appreciate levels of evolutionary control within animalian BrainMinds, we can also better evaluate the levels of consciousness that influence animal and human behavior.

Efforts to make affectively experienced BrainMind changes prematurely “unconscious” have sustained the bias in Western psychology and neuroscience to envision other animals as unconscious agents in the world. But this may be because of a focus on tertiary-process (reflective) rather than primary-process experiences. There are reliable ways to analyze primary-process feelings through animal brain research. The most abundant and theoretically critical data has been generated with localized electrical and chemical brain stimulation, especially of ancient subcortical regions, which generate visually observable emotional patterns. Essentially all such types of stimuli are experienced since they sustain reward and punishment functions within animal BrainMinds. That is the empirical gold standard for affective experiences in animals that cannot speak.

Humans stimulated in these, evolutionarily homologous, brain regions, consistently report powerful affective experiences, consistent with the types of emotional behaviors seen in animals—for example, a system that generates RAGE behaviors in animals generates angry feelings in humans. People experience “ownership” of these feelings: They are one’s own, without feeling they were provoked by external agents. In short, affective anoetic consciousness was aroused without any reasonable autonoetic rationale. These feelings, when sufficiently intense, are never deemed “unconscious” in humans, so similar experiences should not be deemed to be without subjectively experienced contents in animals.
Primary-process affective consciousness in animals is now well established empirically, but even more conceptual clarity is possible: Affective states—positive and negative feelings—can be categorized into distinct varieties. The above focused primarily on subjectively experienced emotional affects arising from the unconditioned emotional behavior systems of animals. But animals also treat various external stimuli as rewards and punishments—yielding many sensory affects (e.g., pleasant and disgusting tastes) that can be studied. We also experience the states of our bodies through internal receptors concentrated in very ancient regions of our brains, engendering feelings such as hunger and thirst. These feelings inform our bodies need for food, water, and many other worldly resources needed for survival. Thus, we should distinguish among the primary-process emotional, sensory, and homeostatic feelings in humans, as well as in other animals.

Can animals also have higher level emotional feelings? Probably. Recent behavioral research from Michael Mendl’s group strongly suggests that animals make higher order affective value interpretations based on general positive and negative affective mood states. This work has yet to be linked to brain research, but hopefully it will be, eventually giving us better models for studying the interactions between various BrainMind levels of control. Thus, it may eventually provide us with solid scientific paths for data and prediction based arguments about cognitive forms of consciousness.

In sum, we are beginning to understand the nature of primal affective-consciousness in many species of mammals. This may provide a lasting causal scientific understanding of consciousness in humans. There appear to be many primary-process ways to feel good and bad about what happens to us in the world. In MindBrain evolution, these raw experiences may have set the stage for the evolution of higher forms of consciousness, which deserve the term awareness. Raw affective experiences have been philosophically, cognitively, and scientifically impenetrable for a long time. However, dual-aspect monism strategies (that core affective mentality emerged from the instinctual emotional action systems of the brain) have finally cut, at least partly, through the Gordian knot. With solid neuroscientific evidence, scientists can welcome many other species into the circle of conscious being. Through their contributions to our research endeavors and understanding the animate world, we can finally begin to specify the brain-body functions from which primary-process experiences arose in evolution.

In sum, solid scientific work, with abundant lines of convergent evidence, now affirms that many other animals are conscious beings. This research is providing guidance for systematic inquiries into the roots of the nested-neural-hierarchies that led to various higher forms of consciousness—in humans, other great apes, and perhaps cetaceans (porpoises, whales) and the like. Whether affective experiences exist in poikilothermic/ectothermic (cold-blooded, in the vernacular) vertebrates and invertebrates must remain an open issue for the time being. Behavioral and pharmacological data collected by investigators such as Robert Huber and colleagues suggests that affectively rewarding processes may exist in the nervous systems of such vertebrates and invertebrates, but the neural and psychological homologies are much harder to specify.

Jaak Panksepp

See also Emotion and Psychopathology; Emotions and Consciousness; Self-Consciousness

Further Readings


Consciousness and Embodiment

For centuries scholars have attempted to understand the nature of human experience (i.e., consciousness). For the past 500 years, at least in Western culture, this issue has revolved primarily around the relationship of consciousness to the body. This is because the “mind-body” question has actually been at center stage of an even larger debate regarding the basic nature of reality. This entry examines how the Western debate regarding consciousness and its relationship to the body emerged historically out of the struggle between religious and scientific approaches to reality. As scholars moved from conceptualizing consciousness in religious terms (i.e., as the soul) and attempted to conceptualize it in scientific terms (e.g., as information), the body’s status as a material object was basically taken for granted. Now, as scholars have increasingly encountered problems with scientific conceptualizations of consciousness, some researchers are reexamining traditional materialist assumptions regarding the body. This reexamination is often referred to as the embodiment movement.

Religious Roots of Scientific Approaches to Consciousness and Embodiment

The beginning of mind-body debate that ultimately gave rise to the current embodiment movement is often traced back to the 16th-century French philosopher René Descartes, who simultaneously believed in the reality of the material and the reality of the spiritual. According to Descartes, while the former was comprised of substance (i.e., matter) entailing spatial and temporal boundaries, the latter (i.e., mind) was infinite and did not possess spatio-temporal properties. According to such a position, what is most often referred to as dualism, human consciousness falls on the spiritual, subjective side of the dichotomy, while the body falls on the material, objective side.

Understanding Descartes’ dualism and his motivations for creating it proves important to an understanding of consciousness and embodiment, for while his commitment to the reality of the spiritual is traceable to his religious convictions, his commitment to the reality of matter stemmed from his scientific convictions. As scholars responded to Descartes’ dualism, they either rejected his account of the spiritual or his account of the material. For example, Baruch (Benedict) de Spinoza, a 17th-century philosopher, rejected the mind-body dialectic at the root of Descartes’ dualism by arguing against Descartes’ notion of substance and against the idea that the mind and the body were constituted of different types of substance. Spinoza argued that in order for something to constitute substance, it (a) could not rely on anything else for its existence, (b) could not be caused by anything else, and (c) could not be reduced to anything else. And the only thing he believed met all three conditions was everything in total. John Locke on the other hand, another 17th-century philosopher, attacked...
Descartes’ assertion that the mind was a spiritual phenomenon. Specifically, Locke argued the mind was a natural phenomenon that entailed all the spatiotemporal properties of any other material object. As a result, Locke felt that the mind was best understood via the scientific method.

The point of discussing different reactions to Descartes’ dualism is not to assess the viability of any of the options per se but to point out that although very few modern scholars agree with dualism, the debate regarding consciousness and embodiment continues to be framed in dualism’s internal-external, subjective-objective framework. For example, since roughly the middle of the 19th century, the Western scholarly community has tended to side with Locke and advocate the “naturalization” of consciousness. To naturalize consciousness means to denounce dualism and claim that consciousness is actually part of material reality. This move toward naturalism was fueled primarily by the success of the sciences, and while it removes the spiritual from scholarly conceptualizations of consciousness, it still entails dualism’s division between internal-external and subjective-objective reality. Instead of conceptualizing consciousness as an internal, spiritual phenomenon, however, contemporary naturalism tends to conceptualize it as an internal, mental function that takes place in the brain and allows one to represent the external world and behave within it.

Twentieth-Century Scientific Approaches to Consciousness and Embodiment

Twentieth-century, scientific approaches to consciousness and embodiment emerged as either commitments or challenges to the internal-external, subjective-objective framework entailed in most naturalist approaches to consciousness. Given the naturalist assertion that consciousness constituted an internal process, research in the United States considered it a poor scientific variable because it could not be directly observed. One could see another person’s brain (if it were surgically exposed), but one could not see another’s dreams, thoughts, or perceptions. As a result, American psychologists developed a science of behavior known as behaviorism in which scientists studied the lawful relationship between external stimuli and observable behavior. Within behaviorism, the body was considered part of the material world, and consciousness was basically ignored.

With the advent of the cognitive revolution following World War II, researchers began talking about internal processes once again. This was due, in large part, to the emergence of computer technology and the mathematics of information processing. In short, via the use of new technologies, it became possible to embed certain mathematical and logical processes into the functioning of computer programs. Cognitive psychologists adopted this as a metaphor for the mind and conceptualized cognitive activity (i.e., mental functions such as perception, thought, and decision making) as information processing algorithms taking place inside the brain. Thus, unlike the behaviorists who basically ignored consciousness, cognitive psychology equated it with internal processing and basically retained the internal-external, subjective-objective distinction that naturalist approaches to consciousness had inherited from Descartes’ dualism.

Over time, information processing technology gave rise to computer systems capable of complex cognitive tasks such as playing chess and recognizing faces and voices. As a result, researchers began to conceptualize mental processes as mental functions whose essence had more to do with “causal relations” than the physical brain. That is, what was important about cognitive dynamics was the causal patterns (i.e., computer programs) they entailed. For example, while mentally translating a word from one language into another, cognitive psychologists assume certain “operations” have to be run on specific “contents” to find the correct answer. In this functionalist approach to mind, the series of operations, as well as the contents on which they operated, were thought of as the formal relations expressed in a mathematical equation or a computer program. Because researchers had been so successful at instantiating these patterns of causal relations (i.e., algorithms) into media other than neural tissue (i.e., in silicon computer chips) scholars conceptualized the mind as being comprised of informational properties whose status as mental operations was independent of the physical properties of the medium in which they occurred. Within this internalist, informational approach to mental phenomena, the body was conceptualized as a physical system that, in the end, was considered irrelevant to cognition and consciousness. As a result, the mind was constituted of information, and the body, of matter. In a sense, Descartes’ notion of “soul” was replaced by the naturalist notion of “information.”
Embodiment Approaches to Consciousness and Embodiment

In the late 20th century, scholars of consciousness began to react to functionalism’s informational account of mental processes, specifically because of its strong division between the mind and the body and its commitment to the idea that the body played no role in explanations of consciousness and mental phenomena. Members of this movement have been loosely referred to as “embodiment” theorists. Different scholars turned to the embodiment movement for different reasons. Some believed the body and the world played a larger role in scientific explanations of mental phenomena than the functionalists had claimed, while others felt that the phenomenological properties of mental phenomena (e.g., what it felt like to perceive and think) could not be sufficiently addressed by referring to a disembodied mind.

Scientific Explanation and Embodiment Theory

Despite the success of disembodied approaches to mind, scientists began to challenge the idea that the mind was best modeled as an internal computer program. For example, embodiment scientists began to argue that high-level cognitive abilities such as language were based more on lower level sensory-motor skills than on formal, propositional structures that functioned like isolated computer programs. Embodiment theorists were led to this conclusion by experiments that indicated access to words is actually constrained by body dynamics (e.g., whether or not one is smiling or frowning while trying to identify a word) and the degree to which a word refers to an action versus a static event.

In robotics, researchers began to create robots equipped with bodies whose dynamics actually constituted part of the problem solving necessary to the robot’s production of complex behavioral sequences. Examples include robots designed to model mate selection behavior in crickets. While an informational view might model mate selection as a logical sequence of operations in which a brain program (a) discriminates between different sounds, (b) determines the direction of the selected sound, and (c) activates a motor program needed to move toward the source, robots were designed to reflect the way that crickets actually solve the problem. Specifically, the cricket’s “ears” are located on its forelegs, and the internal tube connecting them is structured so that only particular frequencies can influence the cricket’s movements. Thus, the bodily composition of the cricket solves what a functionalist approach might regard as “cognitive” problems.

In neuroscience, researchers discovered that brain dynamics do not entail the neat, modular, divisions between perception, action, and cognition typically posed by the informational/computational view. Specifically, it turns out that neural dynamics involved in perception are also involved in planning. That is, certain neurons in the prefrontal, motor-planning areas of the monkey brain become active when a monkey is presented with an object associated with behavior. This implies the monkey perceives the object in terms of the plans (e.g., grasp or pinch) it would use to behave toward the object. In addition to these “canonical” neurons, researchers also discovered neurons that are active during both the planning of a goal-related action and the observation of a goal-related action produced by another. In short, these neurons, which were labeled “mirror” neurons, responded to both the planning and the perception of goal-directed activity. These findings did much to dismantle the notion that cognition works according to clearly segregated processing stages, for if perceiving and planning share overlapping neurodynamics, the two are not clearly discernible functions.

Collectively, these findings support embodiment theory and its assertion that the disembodied, informational approach to cognitive phenomena is scientifically insufficient because it ignores the role the physical body actually plays in cognition.

Consciousness and Embodiment Theory

As embodiment theory was influencing scientific explanations of mental phenomena, other scholars were using it philosophically to address the nature of consciousness. Specifically, some philosophers moved away from conceptualizing consciousness as an informational property of an internal, symbol-processing system, to regarding it as an emergent property of the continuous interactions between brain, body, and world. Thus, for embodiment theorists, consciousness, or “what it feels like,” is heavily contextualized and constrained by the fact the brain is nested within a body, which, in turn, is nested within a particular environmental context.
Some philosophers made this move from computationalism to embodiment theory because computationalism retains the internal-external, subjective-objective divisions of Descartes’ dualism. Such divisions are a problem for theories of consciousness because they beg the question of how one can ever be certain one’s internal processes accurately represent external reality. While Descartes solved this issue by asserting consciousness corresponded correctly to external reality because God had made it so, cognitivists specifically, and naturalists in general, tend to claim the correspondence between internal representations and external reality can be solved via the scientific method. Specifically, correspondence theories assume that multiple observations across multiple researchers in controlled settings will allow us to scientifically determine the intrinsic nature of external reality, independent of observers (i.e., subjectivity). The problem faced by correspondence approaches to truth, as has been stated by many scholars since its inception, is that once one assumes that truth is measured by the degree of correspondence between internal representations and external reality, regardless of whether internal processes are conceptualized as spiritual or material, one has logically denied oneself access to truth. This is because it is logically impossible to compare experience (i.e., representations) to nonexperience (i.e., nonrepresentations). That is, if consciousness is an internal representation, then all that consciousness can ever compare is its own representations to its own representations.

In the attempt to overcome problems faced by correspondence approaches to consciousness, embodiment philosophers work to explain how and why consciousness accompanies brain, body, and world dynamics. For example, some embodiment scholars focus on the content of consciousness (i.e., the particular collection of items one is aware of at any given moment, such as the taste of a scoop of ice cream or the color of a rose) and propose antirepresentational accounts of such content. That is, they argue that the content of consciousness derives not from representations of external reality but from the sensory-motor knowledge embodied in our neuro-muscular architecture. According to these theorists, the conscious experience of the color red is not a neural reaction to a wavelength of energy. Rather, it is the experience of a behavioral possibility, what ecological psychologists refer to as an affordance.

Other embodiment theorists go even further and equate consciousness with life. From this perspective, consciousness is explained in terms of the dynamics that distinguish living systems from nonliving systems. Many such theorists conceptualize living systems as open systems that are able to keep themselves far from thermodynamic equilibrium. What this means is that living systems sustain their integrity by taking in, transforming, and dissipating energy and doing so in a way that leaves them with energy stores available for work (i.e., energy transformation).

The most prominent version of this view proposes that the distinguishing property of living systems is their ability to give rise to all the processes necessary to both sustain and replicate themselves. Such systems are referred to as autopoietic. According to the autopoietic framework, consciousness, what it feels like, is a relational property that exists between an autopoietic system and the processes by which it sustains and replicates itself.

### Naturalism, Correspondence, and Embodiment

While the theory of autopoietic systems and the notion of relational properties seem to constitute a naturalist approach to consciousness and embodiment that avoids the subjective-objective, internal-external divide that cognitivism inherited from Descartes’ dualism, many contemporary philosophers argue that all physicalist theories of consciousness face the problem of epiphenomenalism. That is, describing consciousness in terms of relational properties still does not grant it any causal efficacy, because causality resides in physical properties, not relational properties. Thus, while consciousness might actually exist as a relational property, it plays no causal role in describing reality.

One way to challenge epiphenomenalism is to challenge autopoietic theory’s distinction between physical and relational properties. Scholars who do so propose that physiology itself is relational, in that the dynamics of all living systems are necessarily “about” the contexts in which they have evolved and sustained themselves. As a result, it is unnecessary to divide the body into physiological properties and relational properties. Instead, the body is conceptualized as a self-sustaining embodiment of the multiscale contexts (i.e., phylogenetic, cultural,
social, and developmental) in which it has sustained itself. From this perspective, what was previously conceptualized as external reality is now conceptualized as a self-organizing energy-transformation hierarchy. What this means is that living systems form a hierarchy of systems (i.e., plants, herbivores, and carnivores) that sustain themselves on available energy. For plants, sunlight constitutes the available energy. For herbivores, it’s plants, and for carnivores, it’s herbivores. As systems emerge within this hierarchy, their bodies are “embodiments” of the constraints they must address in order to capture available energy. From this perspective, a lion can be conceptualized as a self-sustaining embodiment of the constraints that need to be addressed to sustain a system on the energy entailed in a zebra. Because such self-sustaining embodiments are necessarily about the contexts they embody, “aboutness” (i.e., meaning) is constitutive of what they are. And consciousness, our phenomenology of “what it feels like,” is an evolved form of embodied aboutness. Conceptualizing consciousness as a form of aboutness prevents one from dividing the body into physical and phenomenal properties, and later trying to find a way to put the phenomenal properties back into the fully described physical system.

While the attempt to conceptualize consciousness in terms of relational properties or aboutness might seem unusual, one could argue its actually represents an attempt to unify consciousness and body in a way that overcomes the problems introduced by the 19th-century turn toward naturalism and its assertion that all statements regarding the constituents of reality must be stated in terms of materiality. In essence, naturalism defined itself negatively as a reaction to dualism: To say something is material is to say it is not spiritual. Embodiment theory, on the other hand, seems to be working toward holism: the belief that all of reality constitutes a unity. From this perspective, all things are inherently interrelated. And it is within this dense web of interrelations that consciousness resides. Not as a by-product of independent physical systems, but as the natural aboutness of reality in general.

The notion of aboutness as a basic constituent of reality was central to a holistic, philosophical movement know as idealism, most well-known via the writings of Georg Wilhelm Friedrich Hegel. Just like naturalism, idealism emerged as a response to dualism. And instead of unifying reality by granting primacy to the material realm and trying to find ways to fit consciousness back into reality, as did the naturalists, the idealists took consciousness to be an historically emergent form of aboutness and conceptualized all of reality as inherently interrelated and, as a result, inherently about. This includes both consciousness and the body. As a result, the mind did not relate to external reality via representations. Rather, the truth of experience was to be gauged by its degree of coherence—that is, the degree to which experiences and statements about experience did not contradict one another.

The real difference between a coherence and a correspondence approach to truth is that the latter promises absolute certainty while the former does not. That is, correspondence asserts the scientific method will allow us to overcome the uncertainty of our internal representations and, in the end, describe reality as it truly is. Coherence theory, on the other hand, conceptualizes all phenomena as contextually dependent and interrelated. As a result, there are no independent, intrinsic properties of things to be known. There can only be statements about that in which we are perpetually embedded. Thus, while coherence theory is compatible with science and allows for tests of the coherence of experience as well as tests of the coherence of statements regarding experience, it does so without the naturalist need for absolute certainty. While correspondence naturalism is currently the dominant Western framework for conceptualizing reality, truth, and science, the advent of embodiment theory might provide a context for coherence frameworks to emerge and sustain themselves. Only time will tell.

J. Scott Jordan

See also Explanatory Gap; Idealism; Mind-Body Problem; Mirror Neurons; Physicalism; Reductive Physicalism

Further Readings
Consciousness and the Unconscious

The distinction between conscious and unconscious mind has long exercised philosophers and psychologists. This entry begins by clarifying what unconscious mind is and then examines what kinds of unconscious mental states human beings have. After noting that some philosophers fundamentally reject the very idea of unconscious mentality, the entry then turns to the application of the conscious/unconscious contrast in theorizing about the mind.

What Is Unconscious Mind?

Unconscious mind is defined by way of contrast with conscious mind. But consciousness has been taken to mean many different things. Historically, consciousness has been viewed as the special kind of knowledge that we have of our own minds. On this view, each of us knows of our own minds in a direct first-person way. Talk of unconscious mind, by way of contrast, is talk of aspects of mind that are not known, or knowable, in this way. Here there is an ambiguity. Sometimes the term unconscious mind denotes a wide range of facts about the mind that we cannot know via introspection: facts about the causes, relations, and conditions of mind. For example, thoughts occur to us without our knowing why. The cause of this mental effect is unconscious—that is, unknown—to the thinker. Such causes are only examples of unconscious mentality in a weak, derivative, sense. They count as mental only insofar as they are causes or conditions of something mental. A much stronger sense of unconscious mind denotes a range of entities—states or events—which themselves meet the conditions for being mental entities but which are not knowable in a direct first-person way. In cognitive psychology, cases of unconscious perception count as mental because they are, in some sense, mental states, not simply because they are causes of mental phenomena. That is, the putative unconscious cognitive states are deemed to be mental insofar as they have some essential or definitional feature of the mind: For example, they are about objects or situations.

The weaker notion of unconscious mentality is unproblematic: We cannot deny that there are unconscious causes and conditions of our mental life. The stronger sense of unconscious mentality is more problematic. From the 17th century to the beginning of the 20th century most philosophers (Gottfried Leibniz was a notable exception) assumed that mentality was essentially bound up with direct first-person self-knowledge. A mental state, on this view, is one that is known or knowable in a direct first-person way. While this conception of mind is consistent with the weaker notion of unconscious mentality, it renders the stronger conception of unconscious mentality incoherent because, on that conception, there are entities that are properly mental but not knowable in this direct first-person way. Rather than viewing mentality in terms of self-knowledge, mentality has come to be viewed, by philosophers and psychologists, in causal and functional terms: It is widely argued that certain kinds of distinctive causal properties of creatures or systems are taken to constitute “aboutness” or intentionality. On this conception of mind the stronger notion of unconscious mentality is unproblematic, and in contemporary philosophy and cognitive science it is typically conscious mentality that is deemed to be problematic.

Unconscious Mind: Psychodynamic and Cognitive Variants

Although the contemporary causal and functional conception of mind renders the stronger notion of unconscious mentality coherent in principle, this
does not, by itself, tell us whether we have any unconscious mental states in this stronger sense, nor does it tell us anything about the kind of unconscious mental states that we have. It is important to note that there have been many different claims made about unconscious mind, and some are more justified than others. This entry cannot expand on the varieties of unconscious mentality in philosophy and psychology, but it can focus on two broad ways of conceiving of unconscious mind and the differences between them.

First, there is the psychodynamic conception of unconscious mind, which includes both unconscious facts about mind (e.g., facts about the nature and origin of emotional conflict) and substantive unconscious mental states (e.g., unconscious wishes for something or attitudes toward something). Whether or not there is good evidence for this kind of unconscious mind is something that has been debated for over a century. Claims about psychodynamic unconscious mind are made by way of explaining sets of behavioral and psychological data, including data about irrational behavior, compulsions, dreams, psychological development, and so on. The argument is that if there are unconscious mental elements, processes, and structures of the kind posited in psychodynamic theories of mind, then we would expect to observe the phenomena that we seek to explain. This kind of explanation is known, following Robert Cummins, as “functional analysis” or “functional explanation.” The psychodynamic theorist argues that for many phenomena, there is no available explanation in terms of conscious mentality or in terms of nonmental phenomena, but by introducing unconscious symbols, representations, and “intentional” states into a functional model, these phenomena can be explained.

Although psychodynamic models of mind are theories of how the human mind works in general, they are applied in giving “interpretations” of individual events or patterns of behavior in a specific subject’s life (a particular dream or an emotional outburst, say). The aim is to explain this unique event—to give an idiographic explanation—in terms of the subject’s unconscious mind, with an appeal to unconscious mental events and processes uniquely molded by specific events in infancy and emotionally salient events throughout life.

The psychodynamic unconscious faces a number of challenges: First, is the model of mind correct about what mental entities there are and about how they relate to one another and to nonmental phenomena? Second, given that there are many different psychodynamic theories of unconscious mind, a second challenge arises: Which, if any, of these competing theories is correct, and are any of them better than competing nonpsychodynamic models of mind? Third, are particular interpretations, as applied to individual subjects, correct? Each of these challenges raises further philosophical questions about the standards by which such claims, theories, and interpretations are to be judged.

Functional explanation is central, not just to psychodynamic models of mind but to cognitive science and cognitive psychology. This brings us on to our second broad conception of unconscious mind: the cognitive unconscious. As with psychodynamic theories, one key goal of cognitive psychology is to explain sets of data about human psychological dispositions and competences. But rather than viewing the mind in terms of dynamic emotion-like flows of mental energy, cognitive science has its roots in a different set of metaphors: that of cognition, computation, and information processing. As with the psychodynamic model, there seem to be phenomena that resist explanation in terms of conscious mentality, where no nonmental explanation is adequate, but the phenomena are explicable if we make appeal to unconscious cognitive states and processes: examples include blindsight, implicit learning and consciousness, implicit memory, visual masking, and subliminal perception. But the cognitive unconscious is quite different to the psychodynamic one. Cognitive unconscious states and events play a role in enabling, linking, or facilitating our conscious and rational psychological life rather than being viewed as a powerful competing psychological agency operative within us. Most of these differences stem from the fact that unconscious mind only comes into view as a projection of particular theories of mind. Not only do psychodynamic theories seek to explain different phenomena from cognitive ones, but they differ with regard to their basic theoretical posits. The issue of what kind of unconscious mind we have thus hinges on the deeper question of which theory of the mind is most defensible.

The Rejection of the Unconscious?
Some contemporary philosophers—John Searle in particular—continue to fly the flag for the once standard view of René Descartes, John Locke,
Franz Brentano, and William James that the stronger notion of unconscious mentality is incoherent. Searle does not simply claim that the definition of mind entails consciousness. His argument is more subtle: Mentality essentially involves intentionality (or *aboutness*). While we can ascribe intentionality to all sorts of things that do not have it (we might say that the parched lawn *wants* a drink) Searle argues that the only “real”—rather than “derived” or “as if”—intentionality is to be found in conscious mental states. Philosophers and psychologists have mistakenly inferred from the fact that it is explanatorily useful to talk *as if* there were unconscious mental states and events, the conclusion that therefore there *really are* unconscious mental states and events. Searle’s critique has considerable importance for our final topic.

**Putting the Conscious/Unconscious Contrast to Use**

It was noted above that consciousness is something that contemporary psychology seeks to explain. The distinction between conscious and unconscious mentality is relevant to this explanatory task. Against the background assumption that consciousness is itself a cognitive process of some kind and that conscious mental states are a species of cognitive state, it seems reasonable to deploy a methodology of “contrastive analysis.” In contrastive analysis the aim is to find closely related types of cognitive state—such as, visual perceptual states—where one type is conscious, the other unconscious. This allows the theorist to pinpoint the distinct contribution that consciousness makes, over and above mere cognition.

Although this contrastive methodology is orthodox in cognitive theorizing about consciousness it faces considerable problems. First, there is the problem that a particular kind of mental state may exhibit more than one kind of consciousness, whereas the closely similar unconscious state may lack more than one kind of consciousness. For example, visual perceptual states are *like* something for the subject (they exhibit *phenomenal consciousness*); they play a distinctive role in reasoning and action (they exhibit access consciousness), and they are available to higher order introspective awareness (“self-consciousness”). Nonconscious visual perceptual states (in blindsight or subliminal perception, say) lack all of these characteristics. Unless examples can be found that allow one to pinpoint the contribution made by each kind of consciousness, the contrastive analysis is limited by an overly broad focus.

The second problem is that contrastive analysis rests on the assumption that cognition is one thing, whereas consciousness is an additional property over and above cognition. But if Searle is right, the contrastive analysis cannot take this form. There is no sense in which consciousness is a feature over and above cognition; consciousness is the *precondition* for real intentionality and cognition. This does not imply that consciousness cannot be explained, just that it is a mistake to try to explain consciousness by making appeal to the distinction between conscious and unconscious mentality.

The third problem is that contrastive cognitive theories of consciousness identify consciousness through the lens of cognitive psychological theory. But there are other ways of drawing the conscious/unconscious contrast. In psychodynamic theory consciousness is not a matter of, say, certain kinds of content being available to executive control centers. Conscious mentality involves fundamentally different kinds of cognitive process, operating according to different principles of mental functioning. Consciousness is not a matter of unconscious cognitive content playing a different role but a matter of radically different kinds of psychological processes.

This brief survey indicates another key contrast between conscious and unconscious mentality. Where conscious states are familiar to all of us in a particularly intimate and direct way throughout our waking lives, unconscious mind is always a product of our epistemic and explanatory activities: What the unconscious mind is, and what use the conscious/unconscious contrast has for theorizing about mind, depends very much on which theory of mind is most acceptable.

*Neil C. Manson*

**See also** Blindsight; Conscious Thinking; Perceptual Consciousness and Attention; Self-Knowledge; Subliminal Perception; Unconscious Perception; Visual Masking

**Further Readings**


Conversation and Dialogue

Conversation is a joint activity in which two or more participants use linguistic forms and nonverbal signals to communicate interactively. Dialogues are conversations between two participants (although the terms dialogue and conversation are often used interchangeably). Face-to-face conversation is universal—engaged in by all human cultures and providing an interactive context in which children learn their native languages. Conversation may also be mediated, such as when electronic technology is used for speech or text. This entry takes an interdisciplinary approach to defining conversation and its key characteristics.

A conversation is not simply a sequence of messages expressed as speaking turns, produced by speakers and received and decoded by addressees. Conversations are structured into adjacency pairs, with first and second parts produced by different speakers as in this example:

**Juliet:** Art thou not Romeo, and a Montague?
**Romeo:** Neither, fair maid, if either thee dislike.

Shakespeare’s *Romeo and Juliet*, Act 2, Scene 2

By itself, Juliet’s utterance does not yet count as a question; she cannot be sure that the man lurking beneath her balcony has heard and understood her until she has the evidence from his response. And Romeo’s answer ends up transforming what might have been left to stand as a yes/no question to something highly relevant to their situation, implicating both its interpersonal and familial risks. Utterances in conversation (whether spoken, typed, or produced manually using sign language) are contingent on one another, such that interpretation depends very much on context.

Real conversation is spontaneous rather than scripted in advance; it is shaped by the coordinated behavior of speakers and addressees. For these reasons, it differs considerably from edited texts. Utterances unfold over time; they are both planned and interpreted incrementally. Consider this excerpt from a telephone conversation between two British acquaintances (*adapted from Svartvik and Quirk’s London-Lund corpus*):

```
Brad: Thanks for ringing
Amanda: right—bye
Brad: bye bye <pause> see you next week
Amanda: see you
Brad: see you soon
Amanda: m (both): <laugh>
Brad: you’re hopeless
Amanda: you’re hopeless
Brad: well <pause> no more than you
Amanda: <laughs> no more than usual either
Brad: no <pause> more <pause> no more than you I said not usual
Amanda: oh I know <pause> I said no more than usual
Brad: ah <pause> I’m sorry
Amanda: have you got a new job yet
<etc.>
```

Compared to Juliet and Romeo’s dialogue, this excerpt seems rather disfluent, but it is actually the artifact of an orderly coordination process. Conversations do not begin and end abruptly but with opening and closing routines with which participants establish that they are willing to begin interacting or are ready to say good-bye. At first, it seems as if Brad is winding down the conversation by initiating a preclosing routine with “thanks for ringing.” This is followed by Amanda’s too-abrupt “bye.” Apparently Brad recognizes that Amanda is not serious about hanging up, so he stays on the line even after responding with “bye-bye.” They proceed to draw things out over the next few turns, culminating in joint laughter that displays mutual awareness of this joint pretense. Amanda’s attempt to chide Brad (“you’re hopeless”) fails,
apparently because Brad doesn’t hear her (or perhaps can’t believe what he is hearing). This leads to a repair sequence, during which Brad requests repetition and Amanda complies. Next, Brad (who by now may have figured out that he is being teased) chides her back with “no more than you,” and Amanda attempts further wordplay (“no more than usual . . .”). But Brad interprets this as her mishearing him. He attempts a hesitant (and unnecessary) repair, to which Amanda disclaims, “Oh I know” and then recaps her failed pun. Brad apologizes awkwardly (after an “ah” displaying his belated recognition of her little joke). Then the intrepid Amanda starts up the conversation again. Although only the participants know what they are thinking moment by moment, even an overhearer can recognize from this transcript that flirting is going on.

Empirical Studies of Conversation

Conversation Analysis

This step-by-step description of how conversation unfolds is inspired by the sociolinguistic approach known as conversation analysis. That approach has been developed by analysts such as Emanuel Schegloff, Gail Jefferson, Harvey Sacks, Charles Goodwin, and John Heritage. Conversation analysts provide detailed accounts of the highly coordinated activities that ensue during naturally occurring, everyday conversations (the interpretation of Brad and Amanda’s call provided here does not begin to do justice to the richness of such accounts). Conversation analysis has traditionally resisted quantitative analysis (sometimes militantly so) but has provided many valuable qualitative insights about the structure and coordination of conversation.

The Experimental Tradition

A contrasting but complementary empirical approach to conversation is the experimental tradition pioneered by Herbert H. Clark and colleagues. Typically, language use by pairs of naive subjects is recorded in a laboratory as they interact in a task assigned by the experimenter. Tasks often involve referential communication in which, in order to do the task, two people must come to believe that they are talking about the same thing; the experimenter observes their task-related actions as evidence of what is being referred to. Consider these excerpts from a referential communication experiment in which participants A and B can hear but not see each other. The task requires them to match duplicates of 12 abstract geometric objects; they do this for the first time in Trial 1 and then match the same objects again in Trials 2 and 3 (with many other objects discussed between these excerpted trials).

Trial 1:

A:  Ah boy this one ah boy alright it looks kinda like, on the right top there’s a square that looks diagonal
B:  uh huh
A:  and you have sort of another like rectangle shape, the . . . like a triangle, angled, and on the bottom it’s ah I don’t know what that is, glass shaped
B:  alright I think I got it
A:  it’s almost like a person kind of in a weird way
B:  yeah like a monk praying or something
A:  right yeah good great
B:  alright I got it

Trial 2:

B:  9 is that monk praying
A:  yup

Trial 3:

A:  number 4 is the monk
B:  ok

The common ground that accumulates as a conversation unfolds (whether in a laboratory experiment or an everyday conversation) enables referring to become more efficient over time.

Entrainment in Conversation

In the previous example, convergence on “the monk” marks that these partners believe they are talking about the same thing and are taking a similar perspective on it. When partners reuse the same forms, this is known as entrainment; another pair discussing the same object in a different conversation is likely to come up with quite a different perspective. For example, 13 pairs in one experiment entrained on 13 distinct expressions for the object in Figure 1.
Conversation and Dialogue

By systematically eliciting evidence such as this, the experimental tradition has established that there is less variability within a conversation than between conversations.

Grounding in Conversation

The previous example illustrates another key process underlying conversation: grounding. Conversation involves more than the exchange of messages in sequences of turns; even utterances about everyday objects can be misunderstood. Because individuals cannot read one another’s minds, they seek and provide evidence from their partners in order to conclude that they understand one another. A single utterance (be it Juliet’s question, Amanda’s attempt at a pun, or A’s initial description in Trial 1) does not by itself constitute a contribution to a conversation; it must be grounded with evidence of how the addressee has understood it or taken it up before it contributes to common ground. Such evidence can be verbal and linguistic, in the form of a relevant reply, or nonverbal or nonlinguistic, as in a nod, a puzzled look, or an “uh-huh.” During face-to-face conversation, visual and nonlinguistic signals known as backchannels are often produced and processed simultaneously with verbal turns. This means that conversing is a massively parallel activity: Speakers and addressees simultaneously plan their own utterances, process what their partner has said, and monitor one another’s reactions. Partners share the responsibility for ensuring that they understand each other well enough for current purposes. They flexibly shift the initiative in the grounding process, providing each other with evidence about their own understanding or seeking such evidence from their partner (i.e., with a clarification question). In Clark’s terminology, linguistically encoded “Track 1” messages pertain to the overt purposes of the conversation, while metalinguistic “Track 2” signals (e.g., back channels) provide evidence for grounding.

Conversations With and Through Computers

Human-computer interaction has sometimes been considered to be a kind of conversation, albeit a constrained and task-oriented one; spoken dialogue systems are computational agents that can interact with people using speech in limited domains such as information retrieval and travel planning. In order to minimize reasoning or speech recognition errors, such systems often seize most or all of the initiative for directing the dialogue.

When people communicate electronically, be it by telephone, e-mail, or texting, communication media enable different Track 2 signals for achieving and coordinating shared meanings. How partners coordinate the grounding process differs depending on whether they are copresent in time and space, as well as on other characteristics of the medium. For example, speaking is usually less effortful than typing, so spoken conversations tend to be wordier than typed conversations. Because typing to a small screen using the thumbs is so cumbersome, texted messages are usually shorter than e-mail messages. People on the phone cannot see one another, so they produce more verbal backchannels than when speaking face-to-face. Speech is ephemeral, whereas text need not be; e-mail utterances can be easily reviewed, edited, and quoted. E-mail does not require senders and recipients to be copresent in time and space, so in that medium conversation unfolds over an extended timescale. E-mail often lands in a mailbox interleaved with other conversations or “threads,” and so e-mail programs facilitate keeping track of conversational context by quoting material from previous messages. Through their distinctive affordances for grounding and the variable costs of using them, communication media play a substantial role in shaping dialogue.
Interdisciplinary Approaches

Task-oriented experiments do not address all of the important joint actions that people do in conversation (such as flirting), but experimental findings have the virtue of being summarizable, replicable, and generalizable. Some researchers have sought to uncover the principles underlying conversation by combining the strengths of both descriptive and hypothesis-testing empirical traditions, via experiments that test predictions developed from sociolinguistic insights. Converging results come from linguistic and computational linguistic analyses of conversational corpora. Finally, innovative head-mounted eye-tracking techniques (e.g., pioneered by Michael Tanenhaus, Daniel Richardson, and colleagues) can now be used to unobtrusively measure speakers’ and addressees’ eye gaze as they spontaneously plan, articulate, interpret, and coordinate their utterances moment-by-moment, online.

Susan E. Brennan

See also Discourse Processing, Models of; Gesture and Language Processing; Multimodal Conversational Systems; Production of Language

Further Readings


Debiasing refers to the wide variety of techniques, methods, and interventions that are designed to eliminate or lessen potential errors, distortions, or mistakes in people’s beliefs, judgments, or decisions. This entry summarizes and describes debiasing by placing techniques into three general categories: (a) motivational, involving strategies such as changing the ways in which incentives or punishments are allocated to decision makers; (b) cognitive, involving strategies such as changing the ways in which decision makers think about a problem; or (c) technological, involving things such as changing the ways in which computers and technologies can be employed to assist problem solving. The entry ends by discussing some further implications for debiasing, such as dual-process models of judgment.

**Debiasing Techniques**

One of the most pervasive qualities of human experience is the capacity to construct a multitude of beliefs, judgments, and decisions. People are certainly skilled and rational enough in their judgments and decisions to get through life perfectly fine in most cases. However, at the same time, they often make predictable and systematic mistakes. As testament to this, an ever-growing number of different biases have been identified and categorized, including biases that affect judgments of self and others, introduce errors into estimates and forecasts, and simply cause us to choose wrongly, exemplified by the heuristics and biases perspective.

When there are problems, people naturally look for solutions. Although debiasing research sometimes appears to be overshadowed by research demonstrating various biases—perhaps it seems more noteworthy to show something is broken rather than how to fix it—both debiasing and biasing are equally important to fully understanding judgment and decision making. Just as the study of biases supplies a roadmap predicting the conditions under which mistakes are likely, the study of debiasing supplies a roadmap describing what may be done about these mistakes. The evidence supporting three general categories of debiasing techniques is extensive and comes from diverse sources. This is illustrated with specific examples.

**Motivational Debiasing**

Motivationally based techniques focus on changing incentives or accountability for accurate decision making. For example, people have a general propensity to try to simplify the world by categorizing things. An object with a flat platform, straight back, and four legs may be characterized as a chair, despite its many variations. But one particularly negative consequence of this tendency is stereotyping. People may similarly characterize others simply because they think that person shares characteristics belonging to a particular group. Although there is debate about the extent to which stereotyping is automatic, incentives such as considering future interactions with a person can sometimes lead to less reliance on stereotypes and more reliance on personalized information. Punishments, such as considering retribution for acting prejudiced, may also lead
people to put greater effort into decisions, resulting in less bias.

Making people feel accountable for judgments is another motivational debiasing technique. For example, people have a general tendency to be overly optimistic, exemplified by research indicating that the majority of people think they are better than average (statistically impossible), think they will complete tasks sooner than they actually do, and think that only good and not bad things will happen to them in the future. However, increasing accountability may help lessen at least some of these biases. For example, if people expect that they will have to explain their reasoning to others, they are more likely to put greater effort into decisions. When preparing to justify decisions to others, people seem better able to anticipate possible flaws in their own reasoning. Thus, going on record, or having to justify one’s logic, seems to make people’s judgments more realistic. Motivations can be social, monetary, or otherwise.

Cognitive Debiasing

Cognitive techniques focus on changing the ways in which people think about or conceptualize a problem. One of the most extensively researched cognitive debiasing techniques requires people to consider the opposite of their initial impressions before making a final decision. In essence, the strategy entails asking, “Are there reasons why my initial judgment may be wrong?” For example, with hindsight bias—the tendency to believe one “knew it all along” after outcomes are known—people are most apt to come up with reasons supporting known outcomes, and thus those outcomes seem inevitable. Thinking about the opposite can work as a debiasing intervention by directing people’s attention to alternative outcomes that might not have otherwise received adequate consideration. This technique seems to work especially well when people can easily think of opposing alternatives.

Metacognitive experiences—people’s thoughts about their thoughts—are also critical to debiasing judgments and decisions. One primary metacognition is consistent with the availability heuristic, which refers to the degree to which people experience thoughts as either easy or difficult to bring to mind. For example, if a person tries to debias hindsight bias by using the consider-the-opposite technique of generating many alternative outcomes but also finds this thought generation subjectively difficult to accomplish, he or she may instead infer that the known outcome is in fact more inevitable, not less (“If it is so hard to generate alternative outcomes then the known outcome must be really inevitable”). In short, the effectiveness of this debiasing technique may backfire. Knowledge of metacognition along with cognition is thus necessary to accurately predict when cognitive debiasing techniques might be successful.

Other cognitive techniques involve education and training. For example, people who have learned the correct rule to calculate the area of a parallelogram simply make fewer errors than those who do not. Similar to mathematics, one presumption is that other rules of judgment might likewise be taught, and these may help to debias decisions. For example, once people have learned that large samples represent a population better than small samples this can be taken into account and potentially lead to more accurate decisions. One limitation of education and training is that it may be difficult to get people to fully appreciate that bias is something that indeed affects them personally, and not just other people. Educational training seems to be most effective when decision rules are concrete and directly applicable.

Technological Debiasing

Technological advances, notably the widespread dispersion of computers, have further increased the potential for debiasing. In fact, many decision tasks are too complex and time consuming to compute without the assistance of technology; for example, consider the complex calculations necessary to launch the space shuttle. Complex decisions are known to be more susceptible to biases and errors. It thus seems logical that, at least superficially, computers can aid complex calculations and help lead to more accurate judgments. Of course, technologies are only as good as the people who use them. The Challenger space shuttle disaster illustrates this. On launch day, decision makers at NASA argued over whether the cold temperatures would be a problem; for example, consider the complex calculations necessary to launch the space shuttle. Complex decisions are known to be more susceptible to biases and errors. It thus seems logical that, at least superficially, computers can aid complex calculations and help lead to more accurate judgments. Of course, technologies are only as good as the people who use them. The Challenger space shuttle disaster illustrates this. On launch day, decision makers at NASA argued over whether the cold temperatures would be a problem; they called up the seven prior launches with O-ring issues and found no association between failure and temperature. However, later on when all 24 previous launches were taken into account, the expanded data revealed a 99% possibility of malfunction.
Technological advances such as various algorithms and decision trees that help people to arrive at particular outcomes can also lead to greater debiasing. These decision aids include checklists for promoting quality decisions, formal models for optimizing choices, and methods for promoting group consensus. The computational acumen necessary to implement these techniques can be relatively simple, or they can be far beyond the mental capacity of any one human being. But even complex equations can now be accurately solved in nanoseconds by computers. A weak link in technological debiasing may lie not only in the human decision makers running computers and writing programs but also in getting people to accept resulting outputs over their own intuitive judgments.

Further Implications

Successful debiasing also requires at least four things. Decision makers must (a) be aware of the potential bias, (b) be motivated to correct the bias, (c) recognize the direction and magnitude of the bias, and (d) be able to adequately control or adjust for the bias. These four things may not always be achievable, but important advances in theory and research continue to point to ways of ultimately reaching that objective. Whether people’s biases can be effectively debiased has very profound implications for virtually all beliefs, judgments, and decisions. Researchers have recently begun to explore several intriguing new directions.

Hard-Wired Biases

One of these directions is the extent to which some biases may be “hard-wired” in human thought processes and are automatic versus effortful. For example, in the visual system, people equate judgments of physical distance with visual clarity. When objects are clear, they are seen as closer. Although normally correct, people can also be tricked by using this heuristic, such as when it is foggy. People often confuse foggy conditions to mean that objects are further away than they actually are. To counteract this, for example, airline pilots are taught to rely more on their instruments than on what they see out of the cockpit window under such conditions (also a good approach for automobile drivers to adopt). In these cases, an automatic heuristic decision process is overridden by a deliberate use of instrumentation to help debias judgments.

Systems 1 and 2

Dual-process models—in which judgments are assumed to occur because of automatic versus effortful processes—can have further implications for debiasing. Human judgment may arise from two general processes. The first process is referred to as System 1; its operations are typically quicker, more automatic, associative, and charged with emotion. This can include perceptual or psychophysically based biases (e.g., foggy weather) or association-based biases (e.g., availability heuristic). The second process is referred to as System 2; its operations are typically slower, more effortful, deliberative, and strategically implemented. The assumption is that System 1 is always operative, and System 2 may kick in when people have sufficient motivation and ability to counteract bias. This does not necessarily imply that all biases result from System 1 and all debiasing results from System 2, as both systems can produce bias or not. However, the potential explanatory power of dual-process models is becoming a growing area of interest and may correspondingly increase an understanding of debiasing.

In conclusion, for the human decision maker, when it comes to debiasing, the glass might be viewed as both half full and half empty. On the one hand, people have many highly useful and adaptive decision-making strategies that allow them to get along perfectly fine most of the time; on the other hand, sometimes these strategies are susceptible to errors, distortions, or mistakes. Several debiasing techniques have been devised to try to eliminate or at least minimize bias. These techniques can be classified as motivational, cognitive, and technological. But rather than being mutually exclusive, the techniques may be best viewed as complementary. Each has its place, and they certainly can be used in conjunction as circumstances warrant. Because people are imperfect and fallible decision makers, no matter which strategies are implemented, it is important to keep in mind that debiasing refers to something that will likely occur to a relative degree rather than completely.

Lawrence J. Sanna

See also Availability Heuristic; Belief and Judgment; Decision Improvement Technologies; Decision Making, Neural Underpinnings; Two System Models of Reasoning
Deception, Linguistic Cues to

Much of human social interaction involves trying to uncover what other people think and feel. At the same time, it is often advantageous for people to try to conceal their thoughts and feelings. This discrepancy has given rise to a substantial body of research on the most effective ways to detect deception. Most experts agree that there is no consistent “tell” that gives away a liar. Instead, the most useful cues to deception are those that reveal something about the liar’s underlying cognitive and emotional state. This entry provides a brief overview of linguistic cues to a deceptive state of mind.

Language contains a variety of cues that may indicate that a person is attempting to be deceptive. Because telling a lie requires the construction and retelling of a false story, true and false stories are thought to be qualitatively different from one another. In general, linguistic approaches to deception tend to focus not on the content of deceptive stories but rather on the way that these stories are told.

The primary purpose of language is to communicate our inner monologue to the outside world. Thus, choosing one pronoun over another may reveal very different ways of thinking about people—we implies a closer relationship than you and I. And in some contexts, using an inappropriate verb tense may reveal attempts to conceal part of the story. In 1994, Susan Smith drowned her children in a lake, fabricating a kidnapping story to cover her crime. According to FBI expert Susan Adams, before Smith was a suspect in the children’s deaths, she told reporters, “My children wanted me. They needed me. And now I can’t help them.” Normally, relatives will speak of a missing person in the present tense. The fact that Smith used the past tense in this context suggested to trained FBI agents that she already viewed them as dead.

Depending on the circumstances, cues to deception may be evident in the structure of a story, the words used to tell it, and the physical characteristics of a person’s speech. In general, the language used to tell deceptive stories has four key features. First, deceptive stories tend to be shorter and more repetitive, presumably as a strategy to avoid contradicting one’s own story. Second, deceptive stories are less complex than true stories. The process of creating a false story consumes cognitive resources, leading liars to tell simpler stories. Third, deceptive stories are less immediate than true stories. Immediate language is direct, clear, relevant, and personal, and is often defined in terms of a general impression by the receiver of the communication. Fourth, deceptive stories tend to contain more negative emotion. Liars may feel guilty, either about lying or about the topic they are discussing. Diary studies of small everyday lies suggest that people feel discomfort and guilt while lying and immediately afterward.

Several studies have focused specifically on usage of different types of words. One useful way to study the words that people use is to count and categorize words using a computer program called Linguistic Inquiry and Word Count (LIWC), developed by James Pennebaker and Martha Francis. LIWC analyzes text (or transcribed speech) on a word-by-word basis. The result of this analysis is a profile of the kinds of words people use. A recent paper by Matt Newman and colleagues used this computer...
program to compare the use of language in true and false stories. Using this approach, Newman and colleagues were able to correctly classify liars and truth tellers at a rate of 61% overall.

Liars’ word usage differed in three main ways, each of which were consistent with the general language patterns described above. First, liars used fewer self-references (e.g., I, me, my), indicative of less immediate language. Second, liars used more negative emotion words (e.g., angry, sad, afraid), consistent with the idea that lying is associated with negative emotion. Third, liars used fewer words that make fine distinctions in a narrative (e.g., but, except, also), suggesting both shorter and less complicated stories.

The key point to studies such as this one is that they offer insight into the deceptive state of mind and open the door to new approaches to lie detection. Rather than getting caught up looking for individual cues, lie catchers can look for signs of cognitive and emotional state. Consistent with this idea, work by Maureen O’Sullivan has found that the most effective lie catchers are those people most attuned to psychological processes. O’Sullivan and her colleagues have identified a small group of “lie detection wizards” who show near-perfect accuracy. These wizards tend to be of above-average intelligence and are highly motivated to catch liars. Most notably, many of the wizards are judges and psychiatrists—occupations that require an in-depth understanding of people and how they use words.

In summary, language offers a unique perspective on people’s thoughts, emotions, and motivations. Research on linguistic cues to deception suggests that what we say is less important than how we say it. Even when talking about the same content, liars will tell the story in a simpler, less immediate, and more negative way than truth tellers. Language is just one of many tools available for detecting deception, but it is a powerful one, with enormous potential.

Matthew L. Newman

See also Emotion, Structural Approaches; Folk Psychology; Guilt; Lie Detection; Production of Language

Further Readings


DECISION IMPROVEMENT TECHNOLOGIES

We use our minds to process information and to make decisions. We make choices individually and in groups. Some decisions are personal and others are organizational. Some decisions meet the intended objectives and produce desired outcomes, but many do not. We build and use decision improvement technologies because research shows that they benefit us in situations with complexity, extensive data, and rapid change. The term decision improvement technologies is very broad and should be interpreted as encompassing any use of computing technology to improve individual, group, or organization decision making.

The terms decision improvement technologies and computerized decision aids refer to a diverse set of tools based on a variety of techniques. Some tools are complex and sophisticated and hence are classified as decision support systems (DSS). Other so-called computerized decision aids and technologies are simple (e.g., a web page with information or a web-based checklist) and hence are not DSS. A decision aid provides assistance or help in reaching a conclusion and making a choice among alternatives. In general, DSS are a subcategory of decision improvement technologies.

This entry briefly discusses the need for supporting decision makers, reviews behavioral decision theory and the history of computerized decision support, and presents a broad framework for
classifying information technology interventions to support decision making.

Need for Decision Support

Our cognitive abilities differ. People encounter complex, multi-causal decision situations that create decision-making challenges. Some of us quickly grasp the facts in a situation and respond in a systematic, intendedly rational way. We try to identify a range of alternatives and make a choice that is consistent with our values and relevant goals and objectives. Some of us respond more intuitively and apply our past experiences to cope with the situation and respond appropriately. Some decision situations create enormous stress for decision makers, and the chosen response is inappropriate or wrong, which leads to more stress. Some decision-making behavior is difficult to explain and understand. If the consequences of a decision are important, then decision makers who are concerned about the quality of their decisions should use any cost-effective computing and information technology to aid, assist, or augment their decision process.

Some situations recur often enough that we develop solution methods to assist people in making the decision. Typically, these solution methods involve using mathematical models or heuristic rules. In other situations, information is provided in a timely way to advise or inform the decision maker(s). Decision processes need to be studied and analyzed to identify the need for a technology intervention and evaluate what technologies would improve decision making.

Builders of decision aids and decision support systems assume their intended users are rational and want to use the technologies. Rationality assumes people use reasoning capabilities and are knowledgeable and logical. Herbert Simon introduced the term “bounded rationality” to describe rational choice behavior that takes into account the limitations of both knowledge and cognitive capacity. Decision improvement technologies attempt to extend the bounds of human rationality.

No computerized system can ensure that a person will act in a rational way when that is not the person’s goal. No decision support system can guarantee a rational solution to a complex problem or guarantee that the correct facts about a situation will be uncovered or that relevant knowledge will be applied to resolve problems. DSS can help us struggle more effectively with the challenge of decision making and planning in a rapidly changing, complex, uncertain, information-rich situation that we have had some experience with resolving. People need to grapple with novel, unique decision situations as best they can.

Decision Support Theory

Research about decision improvement technologies is not based on a single, well-defined theory. Behavioral decision theory research has established that cognitive biases exist. People are predisposed to make choices by the way information is presented and the way analyses are conducted. Debiassing or unbiased presentation has been a secondary motivation for building DSS. Managers accept that some people are biased decision makers, but that does not mean they accept that their own decision making may be biased in a situation where a proposed DSS will be used. In general, cognitive bias has been an issue raised more by academic researchers than by industry consultants and practitioners. If DSS builders are consciously attempting to expand the boundary of rational decision-making behavior, then they must be familiar with the cognitive biases that can impact human information processing. Managers and technologists must explore how decision improvement technologies can reduce significant cognitive biases.

Daniel Kahneman won the Nobel Prize for Economics for his influential work with Amos Tversky on heuristics and biases. Researchers have identified dozens of heuristics and biases. Systematic biases have been found in both individual and group judgment situations. Some biases seem easier to correct using computing technologies than do others.

People generally underestimate how much time is required to complete a specific task. Data can be collected and corrections can be made to adjust time estimates. Decision makers “anchor” on the initial information they receive; that influences how subsequent information is interpreted. Designers of reports can attempt to provide opportunities for decision makers to get more detailed data and view data in various formats like tables and charts. We tend to attribute successes to our own actions and abilities but to attribute failures to bad luck and external factors. Correcting this bias using technology is very difficult. Decision makers tend to offer causal explanations even when the evidence only
suggests correlation. Individuals and groups become committed to a course of action, and decision makers are often constrained by past decisions.

Evidence indicates decision makers do not consistently apply the same decision criteria in similar decision situations. Decision aids can increase decision-making consistency. Decision makers often end their search for evidence quickly and accept the first alternative that is feasible. Good retrieval systems can reduce premature closure of the data gathering process. People tend to remember more recent information and either ignore or forget historical information. Information technology can serve as a convenient memory aid. We tend to gather facts that support our biased conclusions but ignore other facts that might support different conclusions. Data-driven decision support systems can create a shared, single version of the truth about results and actions. When we work in a group with others there is a tendency to avoid conflict and reach a decision with limited critical analysis of facts and ideas. Group support tools have been designed to use anonymity to increase informed dissent and increase brainstorming and idea evaluation.

Herbert A. Simon’s classic book *Administrative Behavior* suggests the following propositions about using decision improvement technologies. First, information stored in computers can increase human rationality if it is easy to access when needed for making decisions. Second, specialization of decision-making functions is largely dependent on developing adequate channels of communication to and from decision makers. Third, if a particular item of knowledge is needed repeatedly in decision making, an organization can anticipate this need and provide the individual with this knowledge prior to the decision to extend the decision maker’s area of rationality. Providing such an item of knowledge is particularly important when there are time limits on making decisions. Information technologies can help direct the attention of decision makers. In summary, computerized decision support increases rationality when it provides relevant, quality information to decision makers when they need it and want it.

**History of Computerized Decision Support**

Some trace the origins of computerized decision support systems to the Lyons Tea Shops business use of a digital computer in 1951. The computer handled the company’s accounts and logistics. Software used weather forecasts to help determine the goods carried by “fresh produce” delivery vans to Lyons’s United Kingdom shops.

On November 4, 1952, for the first time, a computer application was used to assist in predicting the U.S. presidential voting results. The fifth UNIVAC computer built was programmed to analyze the partial results to anticipate the outcome. Early that evening the UNIVAC for the CBS television network election coverage predicted a landslide victory for Dwight D. Eisenhower over Adlai Stevenson. CBS producers did not believe the forecasts and discounted the early results.

A few years later, at the Rand Corporation, George Dantzig worked on a linear programming computer application for optimization problems, and Dick Bellman developed the associated Simplex method of computation and dynamic programming. This research demonstrated the feasibility of using software to calculate results for model-based decision support.

In the mid-1950s, the first large-scale, data-driven DSS was designed by Jay Forrester and George Valley, professors at MIT’s Lincoln Lab. The Semi-Automatic Ground Environment (SAGE) air-defense command and control system was deployed beginning in 1958 and was fully operational in 1963. The name SAGE, a wise mentor, indicated the decision support nature of the system. Some parts of the system remained in operation until 1983. SAGE was designed to coordinate radar stations and to help controllers direct airplanes to intercept incoming hostile aircraft.

In the mid-1960s, Douglas Engelbart and colleagues developed the first hypermedia-groupware system, called NLS (oNline System). NLS had on-screen video teleconferencing and was a forerunner to group decision support systems.

Prior to about 1965, it was very expensive to build large-scale information systems. At this time, the IBM System 360 and other more powerful mainframe and minicomputer systems made it more practical and cost-effective to develop management information systems (MIS) in large companies.

In 1967, Michael S. Scott Morton’s dissertation at Harvard University was a major historical turning point. In 1966, Scott Morton had studied how computers and analytical models could help managers make a recurring key business planning decision. He conducted an experiment in which managers actually used a management decision system (MDS). He built, then implemented and tested, an interactive, model-driven management decision system.
In a 1971 *Sloan Management Review* article, Anthony Gorry and Scott Morton introduced the term decision support system.

**Decision Support Framework**

As technology evolved, new computerized decision support applications were developed. Currently, decision improvement technologies include decision automation systems and communications-driven, data-driven, document-driven, knowledge-driven, and model-driven DSS.

Decision automation refers to computerized systems that make decisions and have some capability to initiate action. Typically decision automation is considered most appropriate for well-structured, clearly defined, routine, or programmed decision situations.

The expanded DSS framework focuses on one major dimension, the dominant component in the architecture that provides functionality. Five types of DSS have been identified: communications-driven, data-driven, document-driven, knowledge-driven, and model-driven. Three secondary dimensions are also identified: targeted users, purpose, and the enabling technology. The framework helps describe and categorize a diverse set of systems. Communications-driven DSS use network and communications technologies to facilitate collaboration, communication, and decision making. Data-driven DSS or business intelligence systems emphasize access to and manipulation of a time-series of internal company data and sometimes external data. Document-driven DSS integrate a variety of storage and processing technologies to provide complete document retrieval and analysis to assist in decision making. Knowledge-driven DSS use knowledge technologies to suggest or recommend actions to managers. Finally, model-driven DSS emphasize access to and manipulation of an algebraic, financial, optimization, or simulation model. Some model-driven DSS use data for forecasts or prediction and are classified as predictive analytics. In general, decision support systems attempt to improve the ability of people to make more timely and better quality decisions.

**Conclusion**

Decision improvement technologies are still changing and improving. DSS can improve the presentation of analyses and reduce decision biases. Designers can reduce inconsistent behavior or inadequate search in situations where those biases will negatively impact decision outcomes. In some group decision situations, group technologies reduce conformity pressures. New mobile technologies are providing opportunities for expanding when and how decision aids and DSS are used.

Information technologies can augment and support the human mind. Decision aids and DSS can improve decision-making outcomes and create other advantageous results. Our challenge is to evaluate when we want to let computers make decisions and when we want to keep humans actively involved in making decisions.

Daniel J. Power

**Author’s Note:** Parts of this entry have been adapted with permission from http://dssresources.com.

**See also** Availability Heuristic; Decision Theory, Philosophical Perspectives; Dissent, Effects on Group Decisions; Group Decision Making; Representativeness Heuristic

**Further Readings**

Decision Support Systems website: http://dssresources.com


**Decision Making, Neural Underpinnings**

Our lives are fraught with decisions. Many involve very simple options, immediate outcomes, and minor
Decision Making, Neural Underpinnings

consequences—as when deciding which way to walk through a crowded market or selecting a side dish to go with our lunch order. Other choices can have life-changing consequences. In which university should I enroll? Do I accept this new job or start my own business? Should I begin the experimental cancer treatment? These choices differ in many ways but share a common process of decision making. In all cases, they require an individual to determine two or more courses of action; to evaluate the likely consequences of those actions, both positive and negative; to weigh different consequences against each other according to some personal preferences; and to choose one action to pursue. Over the past decade, researchers have begun exploring the neural underpinnings of a wide range of decisions, from simple perceptual judgments to realistic economic games. This research is striking in its interdisciplinarity. By applying the techniques of modern neuroscience to key problems in cognitive psychology and behavioral economics, researchers have mapped out brain systems that support many aspects of our decisions. Collectively, this research constitutes the nascent field of “decision neuroscience” or “neuroeconomics.”

Learning About Potential Rewards

Of all aspects of decision making, most well studied has been how organisms determine the value of rewards. It has long been recognized that the motivational properties of rewards come from activity of neurons in key brain regions. Studies in the 1950s found that stimulation of neurons in the rat’s dopamine system had similar effects to delivery of physical rewards. Rats preferentially sought out locations where the stimulation had been delivered and pressed levers to stimulate dopamine neurons even to the point of ignoring food and drink. By the 1990s, single-unit recording in monkeys revealed that dopaminergic neurons in the ventral tegmental area (VTA) responded both to rewards and to information about rewards. Before training, the firing rate of VTA neurons increased whenever a reward (e.g., a squirt of juice) was delivered. After the monkey was trained to expect a reward following a visual cue, however, the same neurons now increased their firing rate to the cue but did not respond to the now-expected rewards (and decreased when an expected reward was omitted).

Subsequent computational models accounted for this pattern of results by claiming that dopamine neurons in the VTA—and its projection targets, such as the nucleus accumbens within the striatum—carry a signal that acts as a reward prediction error (RPE). That is, the firing rate of those neurons increases to unexpected positive information or rewards, remains the same for expected information or rewards, and decreases to unexpected negative information or rewards that are smaller than expected. Support for the RPE hypothesis has come from studies using a wide range of techniques and experimental approaches, from single-unit physiology and direct measurement of dopamine release in animal models, to functional MRI and PET in human participants. Recent work has shown that the dopamine system responds to a wide range of rewards: physical (e.g., juice), economic (e.g., money), perceptual (e.g., attractive faces), and social (e.g., cooperation when playing a game). And interdisciplinary teams have used models of the brain’s reward system to make novel predictions about how to design economic auctions. Important areas of current research involve determining whether the brain’s reward system exhibits any functional specialization, whether

Tracking Decision Variables

A key challenge for identifying the neural basis of decision making lies in the complexity of the choice process. Many important decisions involve a host of competing factors and idiosyncratic preferences. To study decision making in the laboratory, researchers typically create abstract decision problems whose properties can be manipulated experimentally. Accordingly, most experiments involve simple economic choices. Human participants might be asked whether they prefer to receive a guaranteed $20 or to have a 50% chance at winning $50. Similar designs are used in research with animals, who might choose between a small squirt of juice now or a larger squirt of juice in a few seconds. Typical factors manipulated include the magnitude of monetary gains and losses, the probability with which an outcome will occur, and the temporal delay until a reward will be received. By examining how the brain responds to changes in these important decision variables, researchers identify neural systems that support distinct aspects of decision making.
Resolving Uncertainty

In life, most decisions involve uncertainty. Studies of how the brain deals with uncertainty have typically investigated risk—that is, when the probabilities of different outcomes are known or can be estimated. Canonical real-world examples of risky decisions include foraging for food, gambling, or investing, and thus laboratory experiments often mimic these sorts of situations. While much remains to be investigated, converging evidence supports three broad conclusions. First, information about probability modulates neuronal responses at all levels of decision making: from perception and perceptual judgments (e.g., in the visual system), to the evaluation of rewards (e.g., in the dopamine system), through the selection of actions (e.g., in the parietal cortex). Second, several brain regions, including the insular and orbitofrontal cortices, appear to carry signals associated with risk. In neuroimaging studies, activation in these regions tends to increase when a decision involves maximal variance in outcomes. Some recent work suggests that parts of the insular and cingulate cortices may signal changes in risk, such as when a situation goes from being relatively stable and predictable to highly volatile. Third, to deal with uncertainty and make effective decisions, the brain engages regions that support executive control processes. For example, when probabilities are completely unknown—in an extreme form of uncertainty known as ambiguity—there is increased activation in the lateral prefrontal and parietal cortices.

Discounting Rewards Over Time

Adaptive decision making often requires weighing a smaller reward now against a potential larger reward later. Most individuals exhibit “temporal discounting,” in that the value of a reward decays as the delay until its delivery increases. Considerable research has shown that temporal discounting typically does not follow an optimal, continual decay function (i.e., an exponential curve); instead, the value of a reward drops off more steeply over time, as in a hyperbolic function. One influential framework posits that this hyperbolic effect results from the interaction of two brain systems: a myopic, reward-related system that prefers immediacy and discounts very rapidly, and a rational, executive control system that wants to obtain the largest reward and discounts very slowly. Accordingly, whether someone makes an impatient or patient choice depends on the relative activation of these two systems. Other research challenges the two-system perspective by showing that activation within the brain’s reward system tracks the subjective value of a reward, regardless of when it is presented. Recent work from both human neuroimaging and single-unit recording in primates has suggested a potential resolution. In this new hybrid model, the value of an option is determined by computations within a single brain system (i.e., the dopamine system and its projections in prefrontal cortex). However, regions in prefrontal cortex important for self-control—along with perhaps other regions—modulate the value system based on goals, traits, and context. As will be discussed in the following section, a similar sort of hybrid model has been extended to a wider range of decisions.

Integrating Decision Variables in Choice

Under standard economic models, people first compute decision variables of the sort described in the previous section, and then combine those variables to determine the subjective value of each alternative. This sort of model, often called “compensatory decision making,” has two important implications for brain function. First, the brain must convert disparate sorts of information—as when we decide whether to save money, go out for a fancy dinner, or buy a new cell phone—into some sort of common currency for comparison. Second, once each potential outcome’s utility has been converted to a common currency, they must be weighed against each other by some sort of comparator mechanism. Note that neither concept, common currency, nor a comparator mechanism, necessarily reflects the activity of a single neuron nor even of a single brain region. Thus, understanding what decision related computations are implemented in the brain is as important as identifying where those computations occur.

Value Integration in Goal-Directed Choice

Evidence from different methods and experimental designs now suggests that the brain’s ventromedial prefrontal cortex (vmPFC) plays an important role in computing value signals for a wide range of rewards.
Earlier lesion studies revealed that individuals with vmPFC damage exhibit abnormal decision making, even if they have no concomitant impairments in cognitive abilities (e.g., memory, language). These individuals often fail to integrate information correctly in complex decisions, lack insight into the consequences of their actions, and have difficulty learning from prior outcomes. Note that this constellation of deficits often, but not always, results in objectively bad decisions. For example, when faced with a series of choices where risky options have higher value, on average, than safer options, individuals with vmPFC damage may do better than neurologically normal controls because they do not experience regret (and thus become risk averse) following bad outcomes.

Recent work using single-unit recording in non-human primates and neuroimaging in humans has examined how vmPFC responds to different sorts of rewards. Strikingly, the firing rate of many vmPFC neurons increases proportionally to the relative value of potential rewards. If a monkey values grape juice twice as much as water, then a vmPFC neuron will fire twice as frequently when selecting an action that leads to a given quantity of grape juice than one leading to the same quantity of water. Functional neuroimaging studies provide similar evidence for the important role of vmPFC in valuation. Within the past few years, numerous studies have shown that activation within posterior vmPFC tracks decision value, or how much someone would be willing to pay to receive a specific outcome. Notably, this basic conclusion replicates across a wide range of rewards: consumer goods (e.g., inexpensive electronics), snack foods, wines, risky gambles, delayed rewards, and even subjective experiences such as viewing an attractive face. Collectively, these studies indicate that posterior vmPFC may contribute to the comparator mechanism in decision making. An important area of current research, accordingly, is to understand how vmPFC signals are shaped by activity in other brain regions (e.g., information about others’ intentions carried within social cognition regions). Studies using functional connectivity and network analyses will be critical for reaching this goal.

**New Models for Decision Making**

Standard compensatory models are appropriate for decisions such as daily consumer purchases. Yet they are difficult to apply to some real-world decisions, from selecting a college to deciding on a course of medical treatment. To make such decisions, people rarely weigh all possible alternatives and consequences; instead, they simplify the decision problem to a manageable scope by focusing on a few attractive alternatives and their most important consequences. Understanding how the brain deals with such complex decisions presents an ongoing challenge. Some research indicates that cognitive control regions in the medial and lateral prefrontal cortex play important roles in setting up decision problems, perhaps by modulating the striatum, vmPFC, and other reward-related regions. As one example, a region within the dorsomedial prefrontal cortex shows altered activation levels and different functional connectivity to other regions, depending on how a decision problem is represented (e.g., whether someone seeks to maximize gains or minimize losses). Different people may also set up decision problems in very different ways. For example, individuals with varying levels of self-control (e.g., successful vs. failed dieters) show different contributions of lateral prefrontal cortex to decision making. And regions of anterior prefrontal cortex may support exploratory behavior, as when we sacrifice immediate gains to obtain information that improves long-term decisions. New knowledge about frontal cortex function will provide important new insights about these and other decision processes.

**Conclusions**

Much has been learned about the neural mechanisms that support decision making. Important discoveries include how dopaminergic neurons signal reward information, how a range of brain regions represent aspects of risk and ambiguity, how the ventromedial prefrontal cortex supports the integration of value signals during goal-directed decision making, and how connectivity between regions may change based on individual differences and contexts. Over the coming years, new research will extend these and other findings to a larger set of real-world problems. Decision neuroscience will provide increasingly important links between the phenomena studied by economists and psychologists and the new understanding of brain function provided by cognitive neuroscience.

*Scott A. Huettel*
Decision Making and Reward, Computational Perspectives

This entry reviews historical developments in the study of the role of incentives in decision making, emphasizing recent computational approaches that model decision making’s physical basis.

Making a decision means selecting an action from a discrete set of alternatives. The role of reward in decision making has been a focus of interest in psychology at least since the early 1900s, when Edward Thorndike proposed the law of effect to describe how rewards shape animal behavior. Efforts to develop mathematical descriptions of perception and behavior after World War II, for example, led to signal detection theory—which assumes reward-maximizing behavior and uses it to characterize basic perceptual abilities—and to theories of economic preference such as R. Duncan Luce’s choice axiom, in which the probability of one choice over another is independent of the set of third options available. Behaviorism reached the zenith of its influence in the mid-20th century as well, using B. F. Skinner’s automated experimental techniques in an attempt to characterize behavior in terms of its consequences for reinforcement. The result was a valuable body of data and a set of robust behavioral regularities (such as Richard Herrnstein’s “matching law”) that continue to constrain theories today. From a contemporary perspective, though, most theorizing during this period—especially behaviorist theorizing—was noncomputational: That is, it did not involve simulating or mathematically modeling a causal process of reinforcement-guided behavior at any level of physical description, either with a machine or with pen-and-paper calculations. This changed when the postwar computer revolution encouraged researchers to regard cognition as a physical, computational process determined by the state of an organism’s brain.

Reinforcement Learning

Reinforcement learning (RL) has been a powerful force in machine learning, psychology, and neuroscience since the 1980s. It blends a computational approach to decision making with the behaviorist or classical-economic assumption that agents act to maximize, or at least improve, earnings. RL theory developed from the theoretical foundations established in control theory by Richard Bellman in the 1950s and 1960s, but specifically exploited the recursive structure of equations for predicting future reward as a function of an action policy (specifying which action to take in every state of the environment) applied to a discrete state-space representation of the world. (A discrete state-space is a representation consisting of a list of all the unique states in which an agent could find itself; in contrast, a continuous set, such as the set of real numbers used in calculus, is uncountable and could never be exhaustively written down, even in an infinitely long list.) Using the discrete state-space approach, Richard Sutton and Andrew Barto showed that unsupervised, online learning by trial and error was an effective method for creating artificial agents without programming in all possible relevant knowledge—indeed, without the programmer even having this knowledge. Despite this success, RL approaches to decision making are often hindered by their frequent reliance on a compound-serial representation of
time: an extremely memory-intensive representation in which a binary state variable is assigned to every relevant subinterval of a time period and linked in a chain (a binary state variable is a memory slot for a 1 or a 0, indicating whether the world is in the corresponding environmental state).

**Decision Making in Continuous Time**

The study of decision making in continuous time, however, has remained a preeminent concern in psychology, since response time (RT) data are continuous. Psychological decision-making models account for RT data using a variety of real-time mechanisms. Many of these also assume finely “discretized” time (time divided into a large number of small subintervals), but unlike compound-serial models, they do not require a unique representational state variable in memory for every discrete subinterval. For example, in the case of a two-alternative, stimulus-discrimination task (e.g., determining whether a light is bright or dim), an agent might initiate two response processes (requiring one real-valued state variable apiece). Each process samples the stimulus repeatedly, thereby accumulating information and racing the other process to a threshold for producing its corresponding response. Accumulation-to-threshold can be conceived as the tallying of votes in favor of one hypothesis about stimulus identity over another (e.g., bright vs. dim).

When the two tallies compete with each other directly—that is, when a new vote for one hypothesis takes away a vote for the alternative—the difference between tallies is a number that traces out a random walk as it changes over time, much like a stock market price over many days of trading. The resulting model can implement the sequential probability ratio test (SPRT) of statistical decision theory, which chooses between two hypotheses by repeatedly sampling a data source. This is appealing since the SPRT minimizes the average number of stimulus samples needed to achieve any given

---

**Figure 1** Diffusion model of decision making

*Note:* Here, one example of a computer-simulated evidence accumulation process (called a “sample path”) is shown in bold. It begins from a starting point representing prior beliefs about whether the stimulus will be bright or dim. When the stimulus begins, the process is driven downward at a constant rate (on average) by a dim stimulus and upward by a bright stimulus. Noise perturbs the process as it drifts downward. On this trial, a response is made some fixed amount of time after the process crosses the “dim” threshold. The time at which it crosses either threshold defines the model’s *decision time*; the sum of decision time and an additional, small, fixed duration defines the model’s *response time* (RT); and a large number of RTs produces a distribution, illustrated as a separate *RT density* function for each decision threshold.
level of accuracy. It therefore lends itself well to maximizing the expected rate of rewards that could be earned by making a series of correct decisions. It also explains, naturally, why people and animals typically produce a speed-accuracy trade-off: When response thresholds are low in the SPRT (i.e., when thresholds for decisions in favor of each hypothesis are more easily achieved), less time needs to be spent collecting information in order to make a response but the chance of an error increases.

Integrating Different Approaches

Recently, work on integrating RL with continuous-time models of decision making (such as Roger Ratcliff’s diffusion model, Figure 1) has produced novel explanations of basic phenomena in RT data that have resisted any widely accepted explanation. For example, an explanation of faster RTs and increased response probability for more preferred responses comes from models in which the threshold of a random walk process is adapted to maximize the rate of rewards. When this happens, agents can rapidly adapt to a near optimal speed-accuracy trade-off in response to changes in the pace of a task. Threshold-adaptation approaches are complemented by other approaches in which the random walk itself is biased to head in one direction by reward history. Both approaches have lower level counterparts that simulate neural processing in the brain using spiking, integrate-and-fire models.

These encouraging results are still very limited compared to the scope of discrete-time RL. Nevertheless, the nascent integration of discrete and continuous-time approaches may lead to better artificial agent performance and a more thorough understanding of the neural circuits underlying biological agent performance because this integration draws on methods of adaptation suitable for the widest range of physical systems—both systems for which a discrete state space is the best description—and those represented optimally with a continuous state space.

Patrick Simen

Further Readings


Decision Theory, Philosophical Perspectives

Decision theory is concerned with how agents should act when the consequences of their actions are uncertain. The central principle of contemporary decision theory is that the rational choice is the choice that maximizes subjective expected utility. This entry explains what this means and discusses the philosophical motivations and consequences of the theory. The entry will consider some of the main problems and paradoxes that decision theory faces, and some of responses that can be given. Finally, the entry will briefly consider how decision theory applies to choices involving more than one agent.

Decision Theory

The history of decision theory is traced to a gambling puzzle that Chevalier de Méré posed to Blaise Pascal. Here is a simpler version of the same kind of puzzle:

Suppose someone offers you a bet that pays $10 if a fair coin lands Heads but costs you $5 if the fair coin lands Tails. Should you take the bet?

To motivate Pascal’s answer, consider two flawed strategies. First, we might assume the worst is going...
to happen. Then we would make sure the worst result is as good as possible. If the bet is taken, the worst result is that the coin lands Tails, as we would lose $5. If the bet is not taken there is no change in wealth. So if we follow this “maximin” strategy—that is, maximize the minimum payoff—we should not take the bet, as no change is better than losing $5.

Alternatively, we might assume the best result will happen and seek to maximize our gain. If we take the bet, the best result is that the coin lands Heads. This gives us a profit of $10. If the bet is not taken, there is no change in wealth. So if we follow this “maximax” strategy (i.e., maximizing the maximum payoff), we should buy the ticket.

Both of these strategies are flawed because they do not take into account the probability of each outcome. Pascal’s key insight was that the decision should be based on the size of each payoff weighted by the probability that it would occur. This gives us the expected monetary value of each action. As the coin has a probability of ½ of landing either Heads or Tails, taking the bet has an expected monetary value of $10 × ½ – $5 × ½ = $2.5. This is greater than the expected monetary value of not taking the bet, which is 0. Therefore, the rational choice in this case is to buy the ticket. In general, Pascal argued that the rational choice is that which maximizes expected monetary value.

The first important modification to Pascal’s decision theory came from Gabriel Cramer and Daniel Bernoulli. They introduced the concept of utility, which can be thought of as one’s overall level of well-being. Cramer and Bernoulli discovered that money has diminishing marginal utility (see Figure 1). This means that each extra dollar added to your wealth adds a little less to your utility. For example, if a millionaire finds a dollar on the street, it will not make him as happy as if a homeless person had found it. Thus, decisions should not be based on expected monetary value but rather on expected utility. So modified, decision theory says that the rational choice is that which maximizes expected monetary value.

To use the above example, suppose the utilities of winning $10, winning $0, and losing $5 are 100, 0, and –80, respectively. Then the expected utility of taking the bet is 100 × ½ – 80 × ½ = 10. As this is greater than the expected utility of not taking the bet (0), the bet should be taken.

**Representation Theorems**

We need one more modification to get to contemporary decision theory. We saw that Pascal recommended weighting the outcomes by the probability that they occur. But where does this probability come from? The probability is a representation of the agent’s degree of belief that the outcome will occur; it is subjective probability. Thus, we arrive at contemporary decision theory: The rational choice is that which maximizes subjective expected utility.

Subjective probability connects decision theory with the beliefs of the agent. Similarly, utilities connect decision theory with the desires of the agent. An agent’s strong desire for an outcome can be understood in terms of the agent assigning a high utility to that outcome. So we see that decision theory systematizes the way that rational choices are based on beliefs (probabilities) and desires (utilities).

But does decision theory presuppose that rational agents have precise numerical degrees of belief and strength of desires? No. **Representation theorems** show that if an agent’s preferences (choices) satisfy certain constraints, then the agent can be represented as having certain numerical beliefs and desires. These constraints include transitivity, continuity, and independence. Transitivity says that if $a$ is preferred to $b$, and $b$ is preferred to $c$, then $a$ is preferred to $c$. Continuity says that if $a$ is preferred to $b$, and $b$ is preferred to $c$, then there is some gamble between $a$ and $c$ such that the agent is indifferent between the gamble and $b$. Independence says that an agent’s choice between two options should not be affected by an outcome that is independent of the choice (Allais’ paradox below offers an example). The
most influential representation theorems are those of Frank Ramsey, John von Neumann and Oskar Morgenstern, and Leonard Savage and Richard Jeffrey. These differ in their constraints and ontologies, but they all show that agents whose preferences have a sufficiently rich and coherent structure can be represented numerically.

The main import of representation theorems is that they provide the most compelling arguments for the central principle that rational choices maximize subjective expected utility. Representation theorems derive this principle from minimal and intuitively plausible constraints on preferences.

Let us now move on to challenges and applications of decision theory, starting with a direct challenge to the assumption of independence.

**Allais’ Paradox**

Maurice Allais offered the following case as a counterexample to decision theory. Suppose you are offered the following two choices (see Table 1). In Gamble 1, you can choose either Risky1 or Safe1. In Gamble 2, you can choose either Risky2 or Safe2. What would you do in each case?

Decision theory tells us that we should choose either the top (risky) option for both or the bottom (safe) option for both. This is because the only difference between the tables is the third (89%) possibility, and the payoff in this possibility is independent of the choice. The axiom of independence says that your choice should not be affected by payoffs that are independent of that choice. The problem is that many people violate this axiom. Many people choose Safe1 and Risky2 and continue to do so after careful consideration. The reason is that humans are risk averse, leading people to prefer Safe1, where risk can be avoided, and Risky2, where risk cannot be avoided.

One response to this paradox is to maintain that decision theory remains the correct normative theory and risk aversion is confused or irrational. An alternative response is to develop a theory, such as the prospect theory of Daniel Kahneman and Amos Tversky, that matches actual behavior more closely.

**Pascal’s Wager**

A famous early application of decision theory is Pascal’s Wager (see Table 2).

If you believe in God, you get either eternal salvation if God exists or a false belief if he does not. If you do not believe in God, you get either eternal damnation if God exists or a true belief if he does not. As the stakes are so much higher if God exists, Pascal argued that the rational choice is to believe in God as long as there is some positive probability, however small, that God exists.

One objection is that there are many gods one could believe in, whereas Pascal’s argument assumes there is only one. Alternatively, Richard Dawkins points out that God might reward evidence-based belief rather than blind faith in his existence. Third, Alan Hájek argues that if there remains a nonzero chance of eternal salvation even if you don’t believe in God, then there is no benefit to believing in God, as any nonzero number multiplied by infinity is still infinity—that is, \(100 \times \text{infinity} = (1 \times \text{infinity}) = (0.1 \times \text{infinity})\). Hájek’s response assumes that the utility of eternal salvation is infinite. Infinite utilities also lead to other problems, as the next case shows.

**St. Petersburg Paradox**

Suppose a fair coin will be flipped until it lands tails. You will win \(2^n\), where \(n\) is the number of flips.
So if the coin lands tails on the first toss, you win $2. If it lands tails on the second toss, you win $4 and so on. The game is played just once and ends when the coin lands tails. The expected value of this game is infinite (i.e., $\frac{1}{2} \times 2 + \frac{1}{4} \times 4 + \frac{1}{8} \times 8 + \ldots$), so according to decision theory, anyone should be willing to pay any amount of money to play it. But it is clearly not rational to give up everything you own to play this game.

One response to this paradox is to claim that it overlooks the diminishing marginal utility of money. Although this response was historically important for motivating the concept of utility, it is ultimately unsatisfying. The reason is that we can simply replace the dollars won with units of utility. The expected utility of playing the game is now infinite, but it is still not rational to give up everything you own to play.

Another response is to argue that utilities are bounded. In other words, there is a level of utility so high that no additional utility is possible. But given the insatiable appetites of humans, this assumption seems problematic. Moreover, the claim would need to be that a bounded utility function is a constraint on rationality, not just a feature of human psychology.

### Newcomb’s Problem and Causal Decision Theory

Perhaps the most important paradox of recent years is Newcomb’s problem (Table 3). Suppose you are faced with a transparent box that contains $1,000 and an opaque box that contains either $1,000,000 or $0. You may take either both boxes or just the opaque one. The twist is that there is a predictor that you know to have an excellent track record. If this predictor has predicted you will take both boxes, the opaque box is empty. If it has predicted you will take only the opaque box, then that opaque box contains the million dollars. What would you do?

Traditional decision theory (called evidential decision theory in this context) says that you should take only one box, as this maximizes expected utility. But this is considered counterintuitive by many, as the million dollars is already either there or not there at the moment of decision. This worry has led to the development of causal decision theory, which says that you should perform the action that causes the outcomes with the greatest expected utility. Thus, causal decision theory endorses taking both boxes.

<table>
<thead>
<tr>
<th>Predicted “Take Both”</th>
<th>Predicted “Take One”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Take both</td>
<td>$1,000</td>
</tr>
<tr>
<td></td>
<td>$1,001,000</td>
</tr>
<tr>
<td>Take one</td>
<td>$0</td>
</tr>
<tr>
<td></td>
<td>$1,000,000</td>
</tr>
</tbody>
</table>

Newcomb’s problem is a case where the outcome partially depends on the actions of someone (or something) else. The final two problems are more familiar cases where more than one person is involved.

### Game Theory

The outcomes of many of our decisions depend on the decisions of others. This interaction is dramatized by the Prisoners’ Dilemma (see Table 4). Suppose you and a partner are charged with committing a crime together. Interrogated separately, each of you can either stay silent or confess. No matter what the other does, each agent is better off confessing; both confessing is the dominant strategy. It is also a Nash equilibrium, meaning that neither agent has any incentive to change his action, given the action of the other.

<table>
<thead>
<tr>
<th></th>
<th>He Confesses</th>
<th>He Stays Silent</th>
</tr>
</thead>
<tbody>
<tr>
<td>You confess</td>
<td>Each get 5 years</td>
<td>He gets 20 years, you go free</td>
</tr>
<tr>
<td>You stay silent</td>
<td>You get 20 years, he goes free</td>
<td>Both get 6 months</td>
</tr>
</tbody>
</table>

No matter what the other does, each agent is better off confessing; both confessing is the dominant strategy. It is also a Nash equilibrium, meaning that neither agent has any incentive to change his action, given the action of the other.

The philosophical significance of this result is that rational agents combine to generate a suboptimal outcome. Each would be better if they both stayed silent, yet each should confess according to game theory. Some philosophers, such as David Gauthier and Douglas Hofstadter, argue that, in fact, the rational decision is to stay silent, but this remains a minority view.

If the agents face this kind of situation more than once, there is more scope for cooperation. This has led to research among game theorists into the evolution of cooperation and morality.
Declarative/Procedural Model of Language

The basic premise of the declarative/procedural (DP) model or theory is that language critically depends on two long-term memory systems in the brain: declarative and procedural memory. Perhaps most importantly, because the computational, anatomical, physiological, molecular, and genetic substrates of these systems are relatively well studied in both animals and humans, this theoretical approach generates a wide range of well-motivated, specific, and testable predictions about the neurocognition of language that one might have no reason to make based on the study of language alone.

This entry summarizes the two memory systems and their interactions, presents the basic predictions of the model, provides an overview of the evidence, and finally discusses implications and future directions.

Social Choice Theory

Whereas in game theory each agent has its own decision and payoff, in many important cases, such as elections, a group has to make a decision and live with the consequences together. The central question of social choice theory is how individual preferences should be aggregated to result in a group decision.

An important result in social choice theory is Arrow’s Impossibility Result. This states that no voting procedure other than dictatorship can satisfy certain reasonable requirements, for example, that if every voter prefers X over Y, then the group prefers X over Y. One avenue of research this has led to is finding a voting procedure that best satisfies the constraints we think voting procedures should satisfy.

Darren Bradley

See also Allais Paradox; Belief and Judgment; Collective Action; Decision Making, Neural Underpinnings; Dutch Book Arguments; Group Decision Making

Further Readings

Arnauld, A., & Nicole, P. (1964). Logic or the art of thinking. Indianapolis, IN: Bobbs-Merrill. (Original work published 1662)
**Procedural Memory**

This system underlies the implicit (nonconscious) learning of new, as well as the control of already learned, perceptual-motor, and cognitive skills and habits, such as typing, riding a bicycle, or video game playing. It may be specialized, at least in part, for sequences and rules. Learning in the system requires extended practice, though it seems to result in more rapid and automatic processing of skills and knowledge than does learning in declarative memory. Note that the term *procedural memory* is used by the DP model to refer only to one implicit nondeclarative memory system, not to all such systems.

The procedural memory system is composed of a network of interconnected brain structures rooted in frontal/basal ganglia circuits, including premotor cortex and BA 44 in Broca’s area in frontal cortex. Although procedural memory is generally less well understood than declarative memory, evidence suggests that the neurotransmitter dopamine plays an important role in this system, as do the genes for some proteins (e.g., for the proteins FOXP2 and DARPP-32). Other factors may also affect procedural memory, including age (unlike declarative memory, procedural memory seems to be established early in life, after which learning and consolidation in this system may decline).

**Interactions**

The two memory systems interact both cooperatively and competitively in learning and processing. First, the two systems can complement each other in acquiring the same or analogous knowledge, including knowledge of sequences and rules. Declarative memory may acquire knowledge initially, thanks to its rapid acquisition abilities, while the procedural system gradually learns analogous knowledge, which is eventually processed rapidly and automatically. Second, animal and human studies suggest that the two systems also interact competitively, resulting in a seesaw effect. For example, a dysfunction of one system may result in enhanced functioning of the other. Similarly, estrogen seems not only to improve declarative memory but also to suppress procedural memory functionality.

**Predictions of the Model**

According to the DP model, each of the two memory systems is expected to play roles in language that are analogous to those they play in animals and humans in other domains. Declarative memory should underlie all idiosyncratic knowledge in language—that is, the mental lexicon—across linguistic subdomains (e.g., simple words and their meanings, irregular morphology, syntactic complements). Procedural memory should underlie the rule-governed sequencing of complex forms, again across subdomains, including phonology, morphology, and syntax (e.g., *walk* + -ed, *the + cat*). However, complex forms can also be learned and processed in declarative memory for example, as chunks (e.g., *walked, the cat*). Thus, complex forms can rely on either memory system. Which one they rely on will depend on multiple factors, such as which system functions better. Note that the DP model is compatible with the possibility that either memory system may be subspecialized for aspects of language (evolutionarily or emergent through learning) and that additional specialized circuitry for language may exist external to the two systems.

**Evidence**

Overall, converging evidence from multiple lines of experimentation, including behavioral, neurological, developmental, neuroimaging, and electrophysiological studies, suggests the following. Consistent with the basic claims of the DP model, idiosyncratic linguistic knowledge seems to be learned, stored, and processed by declarative memory. Disorders of declarative memory impair this knowledge. For example, the amnesic patient H. M. had trouble learning new words, and patients with damage to portions of temporal neocortex have problems with previously learned words (e.g., in semantic dementia or Alzheimer’s disease) while remaining relatively spared at using rule-governed complex forms.

Rule-governed complex forms often depend on procedural memory. Adult-onset disorders affecting procedural memory brain structures, such as Parkinson’s disease, can impair the use of these forms (e.g., patients might say “Yesterday I walk over there,” leaving off the rule-governed -ed suffix). In contrast, unsuppressed output from the basal ganglia to frontal cortex, as is found in Huntington’s disease and Tourette’s syndrome, can result in the overapplication of rules (e.g., *walkeded, dugged*) or faster processing of complex forms.

Complex forms can also be learned and processed in declarative memory. Which of the two
memory systems they rely on seems to depend on a variety of subject-, task-, and item-level factors. For example, individuals who are better at declarative memory (e.g., women vs. men) or worse at procedural memory (e.g., those with developmental disorders that affect this system, such as specific language impairment or with *FOXP2* mutations) appear to rely more on declarative and less on procedural memory. And consistent with the finding that declarative memory improves during childhood, while procedural memory shows a different pattern, adult second language learners appear to rely more on declarative than procedural memory for complex forms, when such learners are compared to native language speakers.

**Future Directions**

However, much remains to be examined. For example, there has been little work on the endocrine or genetic predictions of the model. Additionally, the model’s pharmacological and pedagogical ramifications may prove important for the rehabilitation of language disorders as well as for second language learning. Future studies will provide a better understanding of the model and its implications.

*Michael T. Ullman*

**See also** Amnesia; Bilingual Language Processing; Gender Differences in Language and Language Use; Memory, Neural Basis

**Further Readings**


**DEDUCTIVE REASONING**

Deductive reasoning is the kind of reasoning in which, roughly, the truth of the input propositions (the premises) logically guarantees the truth of the output proposition (the conclusion), provided that no mistake has been made in the reasoning. The premises may be propositions that the reasoner believes or assumptions that the reasoner is exploring. Deductive reasoning contrasts with inductive reasoning, the kind of reasoning in which the truth of the premises need not guarantee the truth of the conclusion.

For example, a reasoner who infers from the beliefs

1. if the room is dark, then either the light switch is turned off or the bulb has burned out;
2. the room is dark;
3. the light switch is not turned off;

to the conclusion

4. the bulb has burned out

is reasoning deductively. If the three premises are true, the conclusion is guaranteed to be true. By contrast, a reasoner who infers from the belief

1. all swans that have been observed are white;

to the conclusion

2. all swans are white

is reasoning inductively. The premise provides evidential support for the conclusion but does not guarantee its truth. It is compatible with the premise that there is an unobserved black swan.
Deductive reasoning has been intensively studied in cognitive science, psychology, and philosophy. There are many important debates concerning the nature of deductive reasoning. This entry surveys three topics—the relationship between deductive reasoning and logic, the main psychological models of deductive reasoning, and the epistemology of deductive reasoning.

Deductive Reasoning and Logic

Deductive reasoning should be distinguished from logic. Deductive reasoning is a psychological process. In contrast, logic does not describe a psychological process. Logic is the abstract theory of the logical consequence relation, the relation that specifies what follows from what.

While logic does not provide a descriptive theory of reasoning, it is widely accepted that there is a normative connection between logic and reasoning. Yet it has proved difficult to precisely articulate this connection. On a simple proposal, reasoners ought to infer the logical consequences of their beliefs and ought not to hold logically inconsistent beliefs. However, Gilbert Harman and others have provided several arguments against this proposal. For example, in some cases, when reasoners recognize that an implausible claim follows from their beliefs, they should not believe the claim but instead give up one of their antecedent beliefs. Harman further claims that there is no distinctive tie between logic and reasoning. This claim, however, is not widely endorsed. Despite the difficulties of precisely articulating the connection between logic and reasoning, the idea that logic provides a normative standard for reasoning is widely accepted.

The Cognitive Psychology of Deductive Reasoning

Sources of Evidence

The psychological study of deductive reasoning has largely focused on investigating its algorithmic underpinnings. Experimental evidence has come from the performance of subjects, typically undergraduate students, on specific reasoning tasks. In one experimental paradigm, subjects are presented with premises and asked to indicate whether a proposed conclusion follows. In a related paradigm, subjects are presented with premises and asked to generate a conclusion that follows, if one exists.

A notable finding of this research is that subjects are highly prone to errors in evaluating the validity of arguments. For example, a meta-analysis carried out by Walter Schroyens, Walter Schaecken, and Géry D’Ydewalle of 65 studies on conditional reasoning found that while 97% of subjects correctly evaluated modus ponens inferences (if \( p \), then \( q; p \), therefore \( q \)) as valid, only 72% correctly evaluated modus tollens inferences (if \( p \), then \( q; \neg q \), therefore \( \neg p \)) as valid. Worse still, 63% incorrectly identified instances of the fallacy of affirming the consequent (if \( p \), then \( q; q, \neg p \), therefore \( \neg p \)) as valid. This research has also uncovered several content effects that influence subjects’ responses. For example, subjects are more likely to mistakenly identify an invalid argument as valid if its conclusion is believable.

Another important source of evidence comes from the Wason selection task, a paradigm developed by Peter Wason. In a typical experiment, subjects are shown four cards and told that each card has a letter on one side and a numeral on the other. The visible sides of the cards show \( D, K, 3, \) and \( 7 \), respectively. Subjects are presented with the conditional claim: Every card that has a \( D \) on one side has a \( 3 \) on the other side. Subjects are then asked which cards they need to turn over to determine whether the conditional is true. The correct answer is the \( D \) and \( 7 \) cards, since the only way to falsify the conditional is for a card to have a \( D \) on one side without a \( 3 \) on the other. Very few subjects (typically, around 10%) select these two cards.

Interestingly, subjects perform much better on variants of this task. In an experiment carried out by Richard Griggs and James Cox, subjects are told to imagine that they are police officers observing people in a bar. They are shown four cards each representing a person at a table, with the person’s age on one side and the person’s beverage on the other. The visible sides show “drinking a beer,” “drinking a coke,” “16 years of age,” and “22 years of age.” Subjects are presented with the rule: If a person is drinking beer, then the person must be over 19 years of age. Subjects are then asked which cards they need to turn over to determine whether the rule is being violated. Most subjects (74%) correctly select the “drinking a beer” and “16 years of age” cards.

Many variants of the Wason selection task have been investigated. Subjects do better in certain
variants than others. It is controversial how these cases should be distinguished. Apparently, minor changes in wording can have a significant impact. To a first approximation, however, subjects perform better on variants involving more realistic rather than abstract contents and on variants that involve permissions or obligations rather than descriptive conditionals.

The data concerning errors made in deductive reasoning provide the principal way of testing psychological theories of deductive reasoning.

**Mental Logic**

There are two main rival psychological theories of deductive reasoning—the mental logic theory and the mental models theory. Versions of the mental logic theory have been championed by Martin Braine and David O'Brien and by Lance Rips, among others. The central claims of this theory are as follows: Human reasoning makes use of mental representations that resemble the sentences of natural language. In deductive reasoning, reasoners manipulate these representations by applying syntactic rules of inference that resemble the rules of logic.

Versions of the mental logic theory differ over exactly which rules are employed in deductive reasoning. Typically, however, they claim that the rules resemble the rules that appear in natural deduction formulations of formal logic. Some of these rules involve the use of suppositions. For example, the rule of reductio ad absurdum states that if supposing that $p$ leads to an absurdity, then one can conclude that not-$p$.

Mental logic theories explain errors in deductive reasoning tasks by appealing both to the difficulty of applying particular rules and to the need to apply multiple rules in particular tasks. For instance, such theories typically claim that it is more difficult to identify modus tollens as valid (compared to modus ponens) because there is no mental rule corresponding to modus tollens. Reasoners must rely on multiple rules (modus ponens and reductio) to derive its validity. Mental logic theories do not directly explain the effects of content on reasoning. Such effects are typically explained by appeal to additional cognitive processes.

**Mental Models**

The mental models theory has been championed by Philip Johnson-Laird and his collaborators. On this view, deductive reasoning involves diagrammatic rather than language-like representations. In an inference, reasoners construct models that represent the possible states of the world compatible with the premises. They then formulate a putative conclusion that is relevant and informative. Finally, they test this conclusion to make sure that the models do not provide a counterexample to it.

For example, the conditional premise, if $p$ then $q$, can be exhaustively represented with the following three models, each specifying a possible state of the world consistent with the premise:

- $p$ $q$
- not-$p$ $q$
- not-$p$ not-$q$

The mental models theory claims that, to save working memory, reasoners do not typically exhaustively represent the information provided by the premises. An initial mental representation of the conditional might instead be this:

- $[p]$ $q$
  
  The square brackets indicate that the first model is the only model in which $p$ is true. The ellipsis indicates that there may be additional models beyond those explicitly represented.

In a modus ponens inference, the initial mental representation of the premise that if $p$ then $q$ is combined with the representation of the premise that $p$. This eliminates the implicit non-$p$ models represented by the ellipsis. The result is the single model:

- $p$ $q$

Since this is a model in which $q$ is true, there are no models that provide counterexamples to $q$. This yields the information that $q$ follows from the premises.

The mental models theory explains errors in deductive reasoning tasks by appealing to the difficulty of reasoning with multiple models. For example, it is more difficult to identify modus tollens as valid (compared to Modus Ponens) because modus tollens cannot be shown to be valid using the initial mental representation of the conditional premise. Rather, it requires fleshing out the initial mental representation of the conditional premise into the exhaustive representation and combining this with the information that not-$q$. The combination...
eliminates the first two models, leaving the single model:

\[ \neg p \quad \neg q \]

This is a model of \( \neg p \) and so yields the information that \( \neg p \) follows from the premises. This process is more complicated than the process for modus ponens, thus explaining the greater difficulty of identifying modus tollens as valid.

The mental model theory can also explain certain content effects. For example, the bias in favor of believable conclusions is explained by the lack of motivation subjects have to search for counterexamples to believable claims.

**Alternative Views**

While the mental rules and mental models theories have many differences, they both claim that the cognitive mechanism for deductive reasoning is a general-purpose reasoning mechanism central to reasoning and problem solving. There are several alternative views that deny this claim. One view is that humans do not possess a general-purpose mechanism for deductive reasoning but rather a different kind of general-purpose reasoning mechanism—for example, one devoted to probabilistic or explanatory reasoning.

A different view is that humans lack general-purpose reasoning mechanisms (or that such mechanisms are not central to reasoning) but instead employ many special-purpose reasoning mechanisms. This view is often motivated by the performance of subjects in the Wason selection task. The idea is that the greater success of subjects in certain variants of the task is explained by the presence of a special-purpose reasoning mechanism. Patricia Cheng and Keith Holyoak have argued that humans make use of pragmatic reasoning schemas, sets of learned context-sensitive rules tied to particular goals. For example, they argue that we possess rules for reasoning about permissions and obligations. Leda Cosmides has argued that humans possess an innate special-purpose cognitive module for detecting cheating in social exchanges. On both of these views, general-purpose reasoning mechanisms do not play an important role in cognition.

**Dual Process Theories**

A common view among psychologists is that humans have two reasoning systems. System 1 is a relatively fast, cognitively undemanding, automatic reasoning system. System 2 is a relatively slow, cognitively demanding, deliberate reasoning system. Versions of this theory differ in how the two systems should be distinguished and how they interact with one another. The cognitive mechanism for deductive reasoning is presumably a part of System 2.

**The Epistemology of Deductive Reasoning**

A central topic in epistemology concerns the nature of justified belief (roughly, reasonable or rational belief). Competently performed deductive inferences typically preserve justification. For example, if a thinker forms a belief via a modus ponens inference applied to justified beliefs, the resulting belief will typically also be justified. This raises the question of why competent deductive reasoning preserves justification. This is not a skeptical worry about whether justification is preserved but rather an explanatory question about why deductive inference has this positive normative status.

There are several different theories of why deductive reasoning preserves justification. On one view, the explanation is that competent deduction is a reliable cognitive process. If the premises are true, the conclusion is guaranteed also to be true. This is a version of “reliabilism,” a view developed by Alvin Goldman. Reliabilism states that a thinker is justified in holding a belief if it is the output of a cognitive process that either (a) does not depend on any input beliefs and tends to yield truths or (b) depends on input beliefs and tends to yield truths when the input beliefs are true.

Reliabilism faces several problems. For instance, it has difficulty explaining why thinkers who are unreliable through no fault of their own—perhaps because they are being fed misleading experiences by René Descartes’ imagined evil demon—can nevertheless be justified in many of their beliefs. A second problem is that there are reliable deductive inferences that do not preserve justification. For example, inferring the conclusion of a complicated proof directly from its premises would not yield a justified belief. To have a justified belief, one must be aware of the proof. These examples suggest that reliabilism has difficulty capturing the intuitive connection between having a justified belief and being a responsible thinker.

A second view is that a pattern of reasoning only preserves justification if the relevant thinker has an appreciation of the fact that the pattern is truth
Delusions are easier to characterize than to define. A deluded individual has a belief (or beliefs) that is held with a degree of certainty unwarranted by the evidence available to that person, that is strongly resistant to vigorous opposition from others, and that distresses or preoccupies the person or otherwise disrupts his or her everyday functioning. While some aspects of this characterization (e.g., the notion that delusions are beliefs) have been subject to philosophical debate, this entry will focus on issues of classification, etiology, and treatment.

Ordinary Versus Bizarre Delusions

Among other distinctions (e.g., the thematic subtype of delusions; the extent to which delusions are circumscribed; the extent to which delusions are congruous with the patient’s mood), psychiatrists distinguish between bizarre delusions, concerning events and situations beyond the realm of possibility, and so-called ordinary delusions, involving beliefs that could conceivably be true but that are highly implausible given the evidence at hand. As an example of the former, consider the mirrored-self misidentification delusion. Individuals with this delusion believe that the person they see when looking in the mirror is a stranger—a state of affairs clearly at variance with ordinary reality. An example of the latter is the de Clairambault delusion, the conviction that another person—typically someone important or famous—is in love with the deluded individual. Although such a belief may well be implausible, it is nevertheless possible—it could be true.

The ordinary or bizarre distinction has important implications for clinical diagnosis, as the occurrence of bizarre delusions is sufficient for a diagnosis of schizophrenia, while ruling out a diagnosis of delusional disorder. Note in this connection that delusions manifest in a range of both psychiatric and neurological conditions. While they are a hallmark


See also
Analogical Mapping and Reasoning; Thinking;
Two-System Models of Reasoning

Further Readings


Joshua Schechter
of psychotic disorders such as schizophrenia and schizoaffective disorder, they are also observed in cases of dementia, epilepsy, Parkinson’s disease, and traumatic brain injury, among other so-called organic conditions.

Theoretical Accounts
A variety of theoretical accounts of delusions have been proposed. At the cognitive level a general distinction can be made between defense explanations, which conceptualize delusions as purposive constructions serving emotional functions and deficit explanations, which view delusions as stemming from functionless disturbances in ordinary cognitive processes, underpinned by neuroanatomical or neurophysiological abnormalities.

Defense Explanations
Consider the Capgras delusion. Individuals with this delusion believe that a person close to them has been replaced by an impostor. In an early formulation of this condition, Capgras patients were viewed as beset by ambivalent feelings toward the replaced individual—perhaps appropriate feelings of familial love in conflict with unacceptable incestuous desires. The suggestion was that in coming to believe that the object of their desire was actually a fraudulent impersonator of their family member, such individuals had found a way of resolving the tension between their incongruous feelings (sexual desire is not so troubling if the desired individual is merely impersonating a family member).

Richard Bentall, Peter Kinderman, and colleagues are more rigorous contemporary advocates of delusion-as-defense explanations. These authors focus on persecutory delusions, which are the most common delusional subtype. An example would be an individual who believes that his neighbors are trying to harm him by pumping poisonous gas into his house. Bentall, Kinderman, and colleagues have suggested that such delusions serve the defensive function of preserving overt self-esteem. Their formulation is complex, but in essence, the idea is that in attributing unpleasant events to the malicious intent of others, persecutory deluded individuals evade responsibility for their misfortunes and thus avoid activating negative beliefs about themselves.

Deficit Explanations
Hadyn Ellis, Andy Young, and colleagues have provided a deficit explanation of Capgras delusion that contrasts with the motivational formulation outlined above. In their account, Capgras delusion arises when the emotional component of face recognition is disrupted. Faces that are visually familiar are subsequently perceived as emotionally unfamiliar, opening the door to delusional imposter notions. The Capgras delusion is thus conceived not as a means of safeguarding psychic integrity but as the functionless consequence of a neurocognitive defect. This explanation accords with Brendan Maher’s conception of delusions as basically normal responses to unusual perceptual experiences. Various authors have identified candidate perceptual anomalies—underpinned by various neuropsychological abnormalities—potentially associated with a range of delusions. For example, Chris Frith and colleagues have investigated the role of disrupted self-monitoring in the genesis of alien control delusions. Patients with such delusions believe that their actions are under the control of some external agency, perhaps an alien or god. Frith and colleagues have proposed that neuropsychological disruption to an “effference copy”—an internal copy of a motor signal, used to predict the sensory consequences of an action—leads to a loss of the normal experience that one has authored the action in question. The suggestion is that such disruptions underpin the experiences of alien control patients.

Max Coltheart, Robyn Langdon, and colleagues have incorporated the Maherian perspective into their two-factor model of delusions. They agree that an anomalous experience may generate a delusional hypothesis but argue that a second factor is necessary for a delusional hypothesis to be adopted and maintained. This second factor is some kind of abnormality in the mechanisms by which belief candidates are evaluated and is most likely underpinned by damage to the right frontal region of the brain. Delusional individuals, on this view, are not rational evaluators of unusual experiences but are deficient in both experience and reasoning. In this connection consider the proposal of Philippa Garety and colleagues that delusions stem from a tendency to jump to conclusions, whereby delusional individuals amass scant evidence when making probabilistic judgments and form unwarranted conclusions on this basis. This tendency may stem from a propensity to attach inappropriate heightened salience to the contents of consciousness, a propensity that Shitij Kapur has suggested results from dysregulated dopamine transmission leading to the stimulus-independent release of dopamine.
Treatment
Pharmacological treatments have a key role in the management of delusions. The only pharmacological property common to all known antipsychotic drugs is dopamine D2 receptor antagonism, in which the activity of a particular dopamine receptor is blocked or dampened. While these medications alleviate psychosis, their limitations—in particular their side effects—have stimulated the investigation of alternatives to pharmacotherapy. The suggestion that deficient belief evaluation capabilities are involved in the genesis of delusions provides a theoretical justification for treating them with cognitive behavioral therapy, and indeed there is evidence for the efficacy of such therapy.

Ryan McKay

See also Capgras Delusion; Confabulation; Fregoli Delusion; Schizophrenia

Further Readings

Depth Perception
Depth perception is the ability of observers to discriminate the distances of objects, particularly their relative distances, and identify the three-dimensional shape of surfaces. This entry concentrates on sources of visual information contributing to depth perception and separates these into two main types: those that apply when the viewer stands still with one eye closed (pictorial cues) and those that come into play when the viewer moves or opens the other eye.

It is hard to imagine perceiving the world without a perception of depth. Animals move around continually, receiving changing images as they do. If they were unable to relate these images to a stable scene they could not operate successfully in a 3D world. This may explain why clinical disorders of depth perception per se are so rare (as opposed to malfunctions of systems that contribute to depth perception, such as stereoscopic vision).

Generally, in nature, depth vision is a consequence of moving through the world. Most animals have eyes positioned on the side of their head so that they can look out for predators; some have a 360° field of view. They are able to recover depth information about the scene by moving. Binocular stereopsis (seeing depth with two eyes) is not necessary for depth perception and acute stereoscopic vision is almost uniquely the preserve of carnivores. It allows them to remain perfectly still in the grass and yet gain the advantages of depth vision.

Pictorial Cues
Of course, it is possible to perceive depth and the 3D structure of a scene without either moving or using binocular stereopsis. When people look at a photograph, which is taken from a single vantage point, they perceive depth in the scene. There are rare examples when this perception can be thoroughly misleading. For example, Figure 1 shows the Ames room, which appears to be a normal shaped room even when we know, from seeing the people in different parts of the room, that this cannot be the case.

Assumptions about floors being perpendicular to walls and windows or tiles being rectangular are so strong that they affect our perception of depth and size. The Ames room makes clear that we cannot deduce the depth structure of a scene in a photograph unless we make assumptions about the scene. Remarkably often, those assumptions prove to be correct but occasionally they can fail.

Perspective
One pictorial cue is perspective. Parallel lines in the world project to straight lines in the image that eventually come together at a vanishing point, as
shown in Figure 1, for example. This is one of the reasons that the viewer is so convinced by the Ames room: It is highly unlikely that a set of lines in the image would all converge to a point unless the lines in the world were in fact parallel (but, of course, in the Ames room they are not). Over many centuries, artists have discovered how to use perspective in paintings. First, they learned to paint lines that recede toward a single vanishing point (in the 15th century, e.g., Masolino), then much later, they painted lines that recede toward two points (in the 18th century, e.g., Canaletto’s pictures of buildings viewed from one corner), and finally, they drew lines that recede toward three vanishing points (in the 20th century, e.g., M. C. Escher’s pictures). For these perspective cues to be informative, the visual system must assume that converging lines in the image correspond to parallel lines in the real world. This is true sufficiently often for perspective to be a powerful cue to depth.

Other Pictorial Cues

Other pictorial cues include aerial perspective (more distant objects have lower contrast and a bluish tinge), familiar size (the retinal size of a familiar object is smaller when it is further away), interposition (near objects occlude more distant ones), shading (a cue to local surface shape rather than distance), relative height in the image (very distant objects, if attached to the ground, are at the height of the horizon), texture gradients (a cue to surface slant and shape), and so on. When artists paint a picture they use many of these cues to create an
illusion, fooling the viewer into believing that there is depth in the scene when in fact the picture is flat.

**Multiple-View Cues**

A large trompe l’oeil picture can be so realistic that if the viewer shuts one eye and keeps their head still, it can appear extremely realistic. Opening the other eye and moving their head breaks the illusion. Some pavement artists can create a similar effect: Viewed from one place, the picture is compelling, but walk around the picture and the vivid perception of depth disappears. This example, like the Ames room, illustrates the fact that multiple views of a scene give much more information about the 3D structure of a scene than a single view. If the scene were real, with objects at different depths, then the image would change radically as the viewer moves. Objects would be dynamically occluded (move behind one another) as the viewer moves, but this does not happen with a painting. The change in relative position of features in the image is often described as motion parallax. Binocular stereopsis is just a special case of parallax, with the left and right eyes providing two samples from the much larger set of images that the observer receives as they move.

**Breaking Camouflage**

In everyday life, pictorial cues and multiple view cues coexist in images, but the two can be separated. Motion parallax and stereopsis give rise to the perception of depth without there being any pictorial cues present. Figure 2, for example, shows a pair of images that contain no pictorial cues or, at least, any pictorial cues that there are indicate that these are flat surfaces. Yet when the two images are viewed together as a stereo pair or as alternating frames in a movie sequence, a shape emerges, in this case a square. The visual system has broken the camouflage that entirely obscured the object. Camouflage in nature is not usually perfect, as it is here, but it can be very subtle. Stereopsis or motion parallax is a crucial way of overcoming it.

**Feature Matching**

To identify depth in the images shown in Figure 2, the visual system must match up corresponding features in the left and right eyes or across time in the case of the movie sequence. Ultimately, the feature matching must occur at a scale as small as the dots in the Figure 2; otherwise, it would not be possible to identify the sharp boundary between the square and its surround. It is likely, however, that the initial matching up is done between coarser scale features than this, which makes the matching process much simpler. The matching process can then be repeated at progressively finer scales to refine the depth estimate.

**Recovering Object Shape**

Once features have been matched, their depth must be calculated. This can be done in more or less sophisticated ways. For example, working out whether one object is in front of or behind another is easy enough, allowing the observer to detect the square in Figure 2 or to thread a needle. Some tasks require more information than this, for example, the observer may need to know the ratio of depths of features. This is called bas relief depth: For example, observers can recognize characters on the Parthenon frieze despite the fact that the depths of the figures have been squashed. Finally, if the task requires the true shape and size of the object to be known, for example when reaching to pick it up, then more
information is needed, including, for example, an estimate of how far away the object is. Interestingly, judgments of the true shape of objects are often most accurate at about grasping distance.

**Combining Depth Information**

Under normal circumstances, pictorial and multiple-view cues to depth agree with one another and support a common interpretation. We have seen examples in which that is not the case, where multiple-view cues (binocular stereopsis and motion parallax) provide the correct answer but pictorial cues are misleading. Nevertheless, there are many cases in which pictorial cues win out over multiple-view cues when a conflict arises. For example, most viewers of a 3D movie see vivid depth in the scene, more so than in a normal movie, but if they were to turn the glasses round so that the left eye sees what the right eye should see and vice versa, the perceived depth does not reverse. This is different from a stereo pair such as that in Figure 2, where reversing the left and right eye’s images results in a reversal of perceived depth (the square recedes instead of protruding). Rather, in a stereo or 3D movie, the perception when the glasses are reversed is more like that produced by viewing a nonstereo movie. In this situation, pictorial and motion cues dictate the perception, while binocular cues are ignored or vetoed because they are inconsistent with the most likely interpretation of the scene. These examples emphasize that the goal of depth perception, and of perception in general, is to make an informed guess about the nature and layout of the scene.

**Other Senses**

Although all the examples so far have been visual, depth perception is multimodal. The sound produced by an object at different distances changes in systematic ways, not only in overall loudness but also in spectral cues (high frequencies are attenuated more than low frequencies as a sound source moves further away). People are able to distinguish the distance of objects based on these spectral differences. Touch also plays a key role in depth perception, particularly in determining the shape of objects. There is increasing interest in determining the way in which haptic (touch) and visual information is combined to determine the perceived shape of surfaces.

**Choosing Between Interpretations**

Researchers often divide up information about depth into categories, describing multiple depth cues within each sense, such as visual texture, perspective, shading, occlusion, and even subcues such as texture-compression, texture-size-gradient, texture-perspective-convergence and so on. Such subdivisions are, of course, artificial constructs that do not necessarily reflect real distinctions in the way the visual system processes information about depth. If the visual system’s task is to choose between competing hypotheses about the scene, then many different types of information may be relevant, all contributing with a greater or lesser weight to the choice. This is a more tractable problem than deciding how the information from different modules should be combined to form a perceptual representation.

Andrew Glennerster

See also Audition, Neural Basis; Perceptual Constancy; Stereopsis; Visuospatial Reasoning

**Further Readings**


Descriptions

Descriptions are phrases of the form an F, the F, Fs, the Fs, [noun phrase]'s F and [noun phrase]'s Fs (e.g. a woman, the tallest spy, apples, the students, my cat, Peter’s cars). They can be indefinite (e.g., an F, Fs, [noun phrase]'s F, NP’s Fs), definite (e.g., the F, the Fs), singular (e.g., an F, the F, [noun phrase]'s F) or plural (e.g., the Fs, Fs, [noun phrase]'s Fs). In English plural indefinite descriptions lack an article and are for that reason also known as bare plurals.

The semantics of descriptions has been one of the central topics in philosophy for more than a century. Issues at the center of the debate include the following:

- What do description sentences semantically assert? If Taffy is department chair, does “the department chair is tall” assert that Taffy is tall? Does it assert that whoever happens to be the chair is tall? Does it assert something else altogether?

- How do we best account for definite descriptions that seem to pick out a unique object in a scenario in which no unique object satisfies the descriptive component? For example, if there are 10 tables in the room and I say, “The book is on the table,” might I have succeeded in saying something true?

- Do descriptions have referential uses? An expression is used referentially just in case it picks out an object, and the descriptive material (if any) does not fully determine which object is picked out. For example, “The man drinking a martini” functions referentially if it picks out someone who is actually drinking water.

- Are descriptions best treated as quantifiers or predicates? A quantifier expresses quantity (e.g., “every student,” “three men,” “most dogs”). A predicate is an expression that is true of exactly those things that possess the property the predicate expresses. For example, student is true of exactly those things that have the property of being a student. “The Danish spy” functions as a quantifier if it quantifies over Danish spies. “The Danish spy is tall” asserts that the set of Danish spies is a singleton set that is a subset of the set of tall people. “The Danish spy” functions as a predicate if it is true of exactly those things that possess the property of being the (only) Danish spy. If descriptions function as predicates, they cannot occur in argument position but can only occur in predicate position, as in “Soren is the Danish spy.” Predicative descriptions can occur in the predicate restrictor position of a quantifier, as in “An owner of a Porsche is usually smug.” Usually here functions as a quantifier. So the underlying form is this: [Most x: being an owner of a Porsche x](being smug x).

- How do we account for plural and generic descriptions? Can “the students asked questions” be true if only four out of 20 students asked questions? What do “The dinosaur is extinct” and “The Chrysler is sold on the West Coast” assert?

This entry focuses on the contemporary philosophical debate over these issues.

Bertrand Russell’s Theory of Descriptions

Most of the philosophical literature on descriptions has dealt with singular descriptions and has followed Bertrand Russell in treating these phrases on a par with quantified noun phrases such as “some F” and “every F.” In the early 1900s, descriptions were commonly treated as special kinds of proper names. In his well-known book Principles of Mathematics (not to be confused with Principia Mathematica, which Russell later coauthored with Alfred Whitehead), Russell suggested that unlike genuinely proper names, which denote their referents directly, descriptions are to be treated as so-called denoting phrases. Denoting phrases denote their denotation via a denoting concept.

In his seminal work “On Denoting,” published in Mind in 1905, Russell offered a new theory of descriptions. According to the new theory, no phrases denote their denotation indirectly via a denoting concept. Description sentences are equivalent to sentences that do not contain a denoting phrase: “The teacher of Plato is wise,” for example, is equivalent to “There is a unique teacher of Plato, and he is wise.” So even though the description sentence, “The teacher of Plato is wise” is grammatically a subject-predicate sentence, it is not logically a subject-predicate sentence, as it is equivalent to a quantified sentence in which “the teacher of Plato” does not occur. For Russell, descriptions are thus incomplete symbols that have no meaning in
themselves. They can be defined only via a contextual definition and not directly.

Though Russell’s new theory has become something of a paradigm of analytic philosophy, it has been challenged on a number of occasions. In “On Referring” Peter Strawson offered reasons for treating definite descriptions as referential. His main charge against Russell was that Russell incorporates aspects of what a description sentence presupposes into what it says. Intuitively, when the presupposition fails, we have a misfire, not, as Russell would have it, a falsehood. For example, “the king of France is bald” mistakenly presupposes that France is a monarchy, and so an utterance of this sentence has no truth value. As further evidence against Russell, Strawson points out that some definite descriptions denote despite failing to be uniquely satisfied. “The book is on the table” is perfectly felicitous in many contexts, but it is not true in the actual world that there is exactly one book on exactly one table.

Strawson’s Critique

Many attempts have been made to save Russell’s theory from Strawson’s objections. A response to Strawson’s charge that Russell confuses what is presupposed by an uttered sentence with what is said can be found in, for example, Stephen Neale’s book-length treatment of the topic. In recent years the problem of incomplete descriptions has been dealt with, for instance, by appealing to quantifier domain restriction; this strategy has been developed by Jason Stanley and Zoltán Szabó. On their account, “every bottle is on the table” can be true even though not every bottle is on the table, because bottle is thought to be associated with an implicit variable whose value is determined by context. The noun phrase together with the completed variable may denote only, say, the bottles in the kitchen. Stanley has argued that the problem of incomplete descriptions can be resolved in the same way.

Donnellan’s Ambiguity Thesis

Another legendary charge against Russell originated in Keith Donnellan’s “Reference and Definite Descriptions.” Donnellan observed that a description may be used (a) attributively to pick out the object (if any) that uniquely satisfies the description or (b) referentially to denote the object the speaker has in mind, and it may be so used even if the object does not satisfy the description. This is the so-called ambiguity thesis. In one of Donnellan’s examples Jones is on trial, charged with Smith’s murder. According to Donnellan, if we use “Smith’s murderer is insane” intending to pick out Jones, we might well say something true of Jones, even if it should turn out that Smith committed suicide. But if he did, then Russell’s theory predicts that we said something false.

Saul Kripke, Neale, and Peter Ludlow, among others, have since then offered extensive criticism of Donnellan’s theory (and close variants). According to Kripke, the occurrence of referential uses of definite descriptions among English speakers does not disprove that the Russellian analysis is correct for English. Kripke’s argument relies on the notion of pragmatic implication. For example, if I say “the dishes are dirty,” I semantically assert that the dishes are dirty but I may pragmatically implicate that I want you to do them. What is pragmatically implicated by a sentence is not determined by the sentence’s grammar and the conventional meaning of the lexical items. Kripke thought that, by appealing to pragmatics, Russellians can explain the phenomena associated with referential uses without positing a semantic ambiguity. Donnellan’s ambiguity thesis and the pragmatically enriched Russellian account thus accommodate the same phenomena overall, even if they accommodate different phenomena at the semantic level. In addition, the pragmatically enriched Russellian account appeals to pragmatic principles we at any rate have to posit in order to accommodate our intuitions about various other expressions. Kripke concludes, based on this and other reasons, that the pragmatically enriched Russellian account is preferable to Donnellan’s ambiguity thesis.

Neale and Stanley, among others, have later suggested that Donnellan and his critics might both be right. Suppose, for example, that I say (intending to communicate a proposition about a person in my visual field) “The guy is drunk.” It is plausible to think that what I have said is equivalent to “There is a unique guy x, and x = a, and x is drunk” where a is a name of the relevant person. So the statement is used both attributively, to pick out the drunk guy, and referentially, to pick out a.

Descriptions as Predicates

In recent years further challenges have been directed against Russell’s original theory. Delia Graff Fara
has argued that descriptions should be treated as predicates rather than quantifiers. *Is a man* and *is the owner of a Porsche*, as they occur in “Kim is a man” and “Kim is the owner of a Porsche,” are to be interpreted in much the same way as the predicate *is human*. Fara offers two reasons against treating descriptions occurring predicatively as quantifiers. One is that, unlike descriptions, quantifiers do not occur in predicate position (compare the unacceptable “Sam and Lisa are not few students”).

The other is that descriptions do not seem to give rise to the sorts of scope ambiguities they would give rise to if they were quantifiers. Quantifiers may take wide or narrow scope with respect to other operators, as in the following sentence:

1. John didn’t talk to some philosopher.

On the narrow-scope reading, Sentence 1 says that it is not the case that John talked to a philosopher, hence that John didn’t talk to any philosopher; on the wide-scope reading, it says that there is a philosopher that John didn’t talk to. But now consider this sentence:

2. John is not a philosopher.

Russell’s theory predicts that Sentence 2 has a reading according to which there is a philosopher that is not identical to John. But this is not a possible interpretation of Sentence 2, which can only mean that John is not identical to any philosopher. Descriptions in predicative position thus seem to be narrow-scope takers. Prima facie, this is odd if they are quantifiers.

Fara has also offered an argument against Russell’s analysis of descriptions in apparent argument position. Consider

3. An owner of a Porsche is usually smug.

Sentence 3 can be read as saying that some owner of a Porsche is smug most of the time, that Porsche owners, in general, are smug most of the time, or that most Porsche owners are smug. On Russell’s theory, only the first reading is available. On the predicate proposal, the description occurs predicatively, preceded by the adverbial quantifiers *sometimes, generally, or usually*. So all three readings are available.

---

### A Unified Theory

Szabó, Gabriel Segal, and Ludlow have offered a different challenge to Russell’s theory. Consider (from Szabó)

4. A man entered the room with five others. The man took off his hat and gave it to one of the others.

We can imagine Sentence 4 being true. But under Russell’s treatment, it’s contradictory. It is contradictory because under Russell’s treatment, “the man took off his hat” entails that there is exactly one man. Yet the previous sentence entails that there is not exactly one man. Of course, it may be argued that context furnishes an adequate delimitation of the domain of discourse or an adequate completion of the description. However, Szabó thinks this approach is doubtful in cases where the speaker has no particular individual in mind. If Sherlock Holmes deduces Sentence 4 from general clues, he may not know enough to pick out any one of the relevant men. Nonetheless, in uttering Sentence 4 he could be saying something true. This is taken as evidence that sentences containing definite descriptions, such as the analogous sentences containing indefinite descriptions, can be true even if the description is not uniquely satisfied.

### Plural Definite Descriptions

Plural definite descriptions present independent problems. On one Russellian account of plurals, “the Fs are G” implies that every one of the Fs satisfies a singular form of the plural predicate G.Russellians thus mistakenly predict that the following sentences have the same truth conditions:

5. Every one of the students asked questions.

6. The students asked questions.

Berit Brogaard has argued that plural definite descriptions are best treated as having the semantic import of partitives of the form “some of the Fs,” “all of the Fs,” “none of the Fs,” and so on. Partitives tell us how many of the Fs individually or collectively satisfy the predicate. “Every one of the students” tells us that every one of the students satisfies the predicate “is an x such that x asked a question.” In the case of
nonpartitive plural definite descriptions, the force of the quantifier is determined by the speaker's knowledge of the lexical nature of the predicate. In the envisaged context for Sentence 6 the students satisfy the plural predicate “asked questions” just in case some of them satisfy the singular predicate “is an x such that x asked a question.”

Generic Uses

A further criticism of Russell’s theory is that it cannot be extrapolated to account for generic uses of descriptions, as in:

7. The dinosaur is extinct.

Richard Sharvy, Fara, and Brogaard have argued that common nouns such as dinosaur or bear serve sometimes as predicates true of individual animals and sometimes as predicates true of larger taxa, as in “there are two bears in Alaska: the black bear and the grizzly” or “the crustaceans evolved simultaneously.” However, this account does not generalize. Consider

8. The Chrysler is sold on the West Coast.

Sentence 8 doesn’t say that the Chrysler species, as a whole, is sold on the West Coast. It says either that Chryslers, in general, are sold on the West Coast or that some Chryslers are sold on the West Coast. The evidence thus points toward a nonuniform interpretation of generic uses of descriptions.

Berit Brogaard

See also Descriptive Thought; Indexical Thought; Intension and Extension; Object-Dependent Thought

Further Readings


DESCRIPTIVE THOUGHT

In uttering the sentence “the tallest building is less than one year old,” I express a thought—call it α—that is about the tallest building, Burj Khalifa. For me to entertain α there is no requirement that I be acquainted—either through perception, memory, or the reports or representations of others—with Burj Khalifa. It suffices that I grasp the description “the tallest building.” In contrast, the thought I express in uttering “Burj Khalifa is less than one year old”—call it β—appears to be about the building in an entirely different way. To entertain β, acquaintance with Burj Khalifa seems to be required.

Two Species of Thought

Descriptive Thought

The above contrast suggests that there are two ways for a thought to be about a thing (e.g., Burj Khalifa) or kind of thing (e.g., water)—descriptively, as in α, or through acquaintance, as in β.

α has three noteworthy features:

• One can entertain α even if one has no acquaintance with Burj Khalifa.

• α can be true at a possible world w at which the tallest building is not Burj Khalifa. (Following Bertrand Russell’s theory of descriptions, all that is required is that there are buildings at w and that the tallest among these is less than one year old at w.)
• $\alpha$ can exist at possible worlds at which Burj Khalifa does not exist.

**Singular Thought**

Some theorists have held that, in addition to descriptive thought, as exemplified by $\alpha$, there is also a category of singular thought. According to such theorists, entertaining a singular thought requires acquaintance with the object or objects whose properties and relations determine whether or not the thought is true. For example, my thought that Burj Khalifa is less than one year old seems essentially to involve Burj Khalifa. On certain views, it contains Burj Khalifa as a constituent and, moreover, is about Burj Khalifa because of this fact.

Prima facie, this thought, $\beta$, has three noteworthy features:

• One cannot entertain $\beta$ if one is unacquainted with Burj Khalifa.
• $\beta$ is true at a possible world $w$ just in case Burj Khalifa exists at $w$ and is less than one year old at $w$.
• The identity conditions for $\beta$ necessarily involve Burj Khalifa—$\beta$ cannot exist at a possible world at which Burj Khalifa does not exist.

Descriptive thoughts are object independent; singular thoughts, object dependent.

**Doubts About Singular Thoughts**

No one doubts that there are object-independent thoughts, but the very idea that object-dependent thoughts exist has been controversial. If, as many have supposed, entertaining a thought involving the concept water requires acquaintance with water, then it seems that one can know that water exists merely on the basis of the fact that one has entertained such thoughts. But it seems doubtful that factual knowledge of this sort can be gained simply by reflecting on one’s concepts. Another major concern has been whether the singular-thought theorist can adequately explain how a thought such as $\beta$ is about Burj Khalifa, as opposed to some other entity or no entity at all. To claim that $\beta$ is about Burj Khalifa simply in virtue of the fact that it contains Burj Khalifa as a constituent merely raises the question, how it comes to contain this entity. Some theorists have sought to analyze this containment in terms of a causal connection between $\beta$ and Burj Khalifa. However, it has proven difficult to make this idea precise, and counterexamples to the general approach have been produced (e.g., by Simon Blackburn). In contrast, if we view thought as uniformly descriptive, we seem to have a straightforward account of aboutness: A thought is about $x$ solely in virtue of its containing descriptive information that $x$ uniquely satisfies. Another advantage to holding that thoughts are uniformly descriptive is that it offers an explanation of how Lois Lane can simultaneously believe and disbelieve, of Superman (i.e., Clark Kent), that he is a superhero: Lois can believe Superman to be a superhero under one description and disbelieve this under another description. The singular thought theorist would seem to be at a loss to explain this since, on her view, the thought that Superman is a superhero just is the thought that Clark Kent is a superhero. (Some singular thought theorists invoke object-dependent, nondescriptive thought components to explain Lois’s cognitive situation. But the question we considered above arises here as well: One might still wonder what makes it the case that one of these object-dependent thought components comes to contain Superman, as opposed to some other entity or no entity at all.)

In light of these considerations, it has been suggested that the appearance of object-dependency in such cases is illusory and that all thought is descriptive; let’s take a closer look at this view.

**Descriptivist Treatments of Singular Thought**

**Ambitious Descriptivism**

Let’s use the term “descriptivism” to designate the view that all thought is in fact descriptive—that, appearances to the contrary, there are no thoughts possessing the characteristics ascribed to $\beta$ above. On this view, the thought that Burj Khalifa is less than one year old is in fact a descriptive thought of the form: The $F$ is less than one year old. On the most ambitious version of this approach, the description the $F$ cannot make reference to individuals. Otherwise, we merely trade one object-dependent (or de re) thought for another; this would hardly mark a genuine advance over the original singular thought theory. (For example, claiming that $\beta$ is in reality the thought that the most famous building in Dubai is less than one year old simply replaces a
thought that is object-dependent with respect to Burj Khalifa with a thought that is object-dependent with respect to Dubai; no progress has been made.) But this ambitious form of descriptivism seems to raise the bar too high. As P. F. Strawson has argued, it will, in almost every case, be impossible to replace what appears to be a reference to a particular individual with purely descriptive conditions and retain the same thought or even a truth-conditionally equivalent thought—that is, one that is true at exactly the same circumstances. (This is especially clear in the case of demonstrative thought.) In addition, even if it turns out that the strategy works in some cases, this will not show or make in any way more plausible that it will always do so.

**A Russellian Alternative**

While the prospects for an ambitious descriptivism appear bleak, a natural intermediate position suggests itself, one initially proposed by Bertrand Russell and subsequently elaborated on by Stephen Schiffer. On this view, $\beta$ in fact exemplifies a certain species of descriptive thought, one in which the only “object” the relevant description makes reference to is the self (and, perhaps, the current time). Thus, in contrast to the suggestion considered above—that $\beta$ is just the descriptive thought *that the most famous building in Dubai is less than one year old*—the current proposal would provide the following: The building that *bears R to me* is less than one year old. In this context $R$ might stand for the relation that holds between $x$ and $y$ just in case $x$ is the most salient object currently in $y$’s visual field (note that there is no requirement that $R$ stand for the same relation in every context).

Why is reference to the self (or the current time) permitted? The answer can be seen most clearly if we consider demonstrative thoughts, such as the thought expressed by my utterance of “that man is smoking.” The description this thought contains must involve reference to me, either directly or indirectly, through reference to some item in the context of my utterance (which is, after all, identified only by reference to the speaker). Of course, this means only that self-ascriptive—or *de se*—singular thoughts are unavoidable; it doesn’t show them to be unproblematic.

But *de se* thoughts are unproblematic in at least two respects in which garden-variety singular thoughts are problematic.

- Setting aside skeptical worries, the fact that I have *de se* thoughts is excellent evidence that the self exists (even though having *water* thoughts is less than excellent evidence that water exists).
- While it is possible for a rational agent to adopt conflicting attitudes to one and the same garden-variety singular thought, this is not so in the case of *de se* thoughts. For example, while *Lois* can simultaneously both believe and disbelieve, *of Superman*, that he is a superhero, this is not possible for Superman himself, at least not in the special case where his thought is *de se*. That is, Superman cannot simultaneously both believe and disbelieve, *of himself*, that he is a superhero. (Similar remarks apply to attitudes involving indexical reference to the current time.)

**Problems With the Alternative Approach**

But even the approach advocated by Russell and Schiffer appears too strong. Perhaps I’m looking at a tree with thousands of indistinguishable leaves. Focusing on one, I think to myself, *that’s green*. The antidescriptivist denies that there is a description that counts as a plausible candidate for the relevant thought component. In particular, he claims that there is no relation, $R$, such that the content of the italicized sentence is that the $x$ that bears $R$ to me is green. This is shown, he claims, by considering the following: For every available value of $R$, I can doubt whether *that one* is the $x$ that bears $R$ to me.

The Russell-inspired descriptivist responds that there is an obvious candidate for $R$—namely, *the x such that x is a leaf that I currently have my gaze fixed on*. Surely, I cannot coherently doubt that *that one* is the leaf I currently have my gaze fixed on. This, however, changes the subject, failing to capture the content of the thought I recognize myself to have; at best, it provides a truth-conditionally equivalent thought (though this is itself doubtful). Moreover, some individuals capable of thinking the original thought may lack the conceptual sophistication required to think the suggested descriptive counterpart.

Finally, there is a more general worry that descriptivism is too liberal—that there is more to thinking about $x$ than entertaining a thought whose description picks out $x$ uniquely. Consider Smith’s belief that the shortest spy is less than 6 feet tall. While it is uncontroversial that there is some individual
whose being less than 6 feet tall makes this belief true, it doesn’t follow that Smith believes, of that individual, that she’s less than 6 feet tall—after all, Smith may have this belief without himself being acquainted with any spies. The descriptivist might acknowledge that apparently singular thoughts require more than just descriptive accuracy—that, for Smith’s thought to be about the shortest spy, the descriptive component must bear a specific (albeit yet to be made precise) causal connection to that very individual. But then one crucial advantage of descriptivism is lost—its ability to circumvent messy questions about acquaintance or causal connectedness. The descriptivist has other debts as well. While the singular thought theorist is obliged to explain how a thought can have an individual as a constituent, the descriptivist must explain how a thought can have a property as a constituent. In addition, while the descriptivist has the resources to explain how Lois Lane can rationally believe and disbelieve the same apparently singular thought—that Superman (i.e., Clark Kent) is a superhero—she lacks the resources to explain a closely related phenomenon: How it is that Bert can rationally believe that dogs make good pets and at the same time disbelieve that canines make good pets. If the property of being a dog just is the property of being a canine, then Bert can simultaneously believe and disbelieve that dogs (i.e., canines) make good pets. This is an intolerable result, one that the descriptivist has no special way of handling.

Gary Ostertag

See also Descriptions; Indexical Thought; Knowledge by Acquaintance; Object-Dependent Thought

Further Readings


**Desirable Difficulties Perspective on Learning**

Instructors and students alike are susceptible to assuming that conditions of instruction that enhance performance during instruction are conditions that also enhance long-term learning. That assumption, however, is sometimes dramatically wrong: Manipulations that speed the apparent rate of acquisition of knowledge and skills during instruction can fail to support both their long-term retention and their transfer to new settings, whereas other manipulations that introduce difficulties and slow the apparent rate of acquisition can enhance post-instruction recall and transfer. Such manipulations, labeled desirably difficulties by Robert Bjork, include spacing rather than massing repeated study opportunities; interleaving rather than blocking practice on separate topics; varying how to-be-learned material is presented; providing intermittent, rather than continuous, feedback; and using tests, rather than presentations, as learning events. That learning profits from contending with such difficulties provides a valuable perspective on how humans learn.

**Learning Versus Performance**

Basically, current performance, which is something we can observe, is an unreliable index of learning, which we must infer. The distinction between learning and performance goes back to research carried out during the 1930s and into the 1950s—research that demonstrated that considerable learning could take place across periods when there were no systematic changes in performance. Experiments on latent learning, for example, showed that rats, after a period of wandering, apparently aimlessly, in a maze, exhibited considerable learning once some target behavior, such as finding a baited goal box, was reinforced. Similarly, human and animal experiments on overlearning—that is, providing additional learning trials after performance had reached an asymptotic level and was no longer changing—demonstrated that such trials continued to enhance learning, as measured by reduced forgetting or accelerated relearning. More recently, a variety of human memory experiments have shown the converse is true as well: Substantial changes in performance can be accompanied by little or no learning. Massed
practice on a task, for example, often leads to rapid gains in performance but little or no effect on learning, as measured by long-term retention or transfer.

**Perspective on Learning**

That certain difficulties can enhance learning highlights some unique characteristics of how humans learn and remember or fail to learn and remember. We do not store information, for example, by making any kind of literal copy of that information. Rather, we encode and store new information by relating it to what we already know—that is, by mapping it onto, and linking it up with, information that already exists in our memories. New information is stored in terms of its meaning to us, as defined by its relationships to other information in our memories. Storing information, rather than using up memory capacity, creates opportunities for additional storage.

The retrieval processes that characterize human memory are unique, too, and differ markedly from a playback of the type that might characterize a typical recording device. Retrieval of information is inferential and reconstructive rather than literal, and it is also fallible, in part because what is accessible from memory is heavily dependent on current cues, including environmental, interpersonal, body-state, and mood-state cues. In addition, and importantly, the act of retrieving information is itself a potent learning event. Retrieved information, rather than being left in the same state it was in prior to being recalled, becomes more recallable in the future than it would have been otherwise, and competing information associated with the same cues can become less recallable in the future. Using our memories, in effect, alters our memories.

From an instruction standpoint, the goal is to create conditions that foster storage and enhance later retrieval, not just at a delay, but also in multiple contexts. We want, in short, to create durable and flexible access to to-be-learned information and procedures. Toward achieving that goal, the conditions of learning need to induce encoding and retrieval processes that are substantial and varied, and incorporating desirable difficulties can induce those processes. Using tests or generation activities—versus presentations—as learning opportunities, for example, exercises retrieval processes that will be needed later. Varying the context, examples, and problem types can lead to a more elaborated encoding of concepts and ideas, which will support transfer of that learning to new settings. Interleaving, rather than blocking, instruction on related topics can aid in encoding the higher order relationships that capture similarities and differences among those topics.

It goes beyond the scope of this entry to provide a detailed discussion of how each desirable difficulty might enrich encoding and/or retrieval processes, but one common characteristic is that such manipulations are likely to induce more *transfer-appropriate processing*—that is, processes of the types that will be required in postinstruction environments.

**The Word Desirable Is Important**

Finally, it is necessary to emphasize that many of the difficulties that can be created for learners are *not* desirable. Desirable difficulties are desirable because they encourage encoding and/or retrieval activities that support learning. To the extent, in fact, that a given learner is not equipped to overcome a difficulty that would otherwise be desirable, it becomes an undesirable difficulty. Thus, for example, requiring that some information or procedure be generated, rather than simply presenting that information or procedure, introduces a desirable difficulty, one that can foster a learner’s subsequent ability to produce that information or procedure, but the learner must be equipped, by virtue of prior knowledge and current cues, to succeed in the generation—or it becomes an undesirable difficulty.

Robert Allen Bjork

See also Discrimination Learning, Training Methods; Memory Recall, Dynamics; Retrieval Practice (Testing) Effect; Skill Learning, Enhancement of; Spacing Effect; Spacing Effect, Practical Applications

**Further Readings**


Desire is a state of mind. Known also as wanting and wishing, this state of mind is one that is a normal cause of coordinated actions, feelings, and thought processes. For example, a desire that one’s father be healthy is a normal cause of reminding him to take his medication (and other actions), a normal cause of feeling disturbed when one learns he has developed a cataract and feeling relieved when the cataract is successfully treated (and other feelings), and a normal cause of paying attention to news items about health care for older men (and other thought processes).

Although desires are commonly associated with actions, feelings, and thoughts, there remain two key questions about desires for theorists to answer. First, which of these effects (if any) are essential to desires and which are inessential? Second, which of these effects (if any) have normal causes other than desires?

Desires and related notions (appetites, passions, and the like) have been discussed by philosophers since at least ancient Greece. It has only been since the 1980s that advances in neuroscience have made empirical science a substantial partner with philosophy in the investigation of desire, but already the payoffs have been substantial.

Philosophical Theories

Philosophical theories of desire come in three main types. One holds that the essence of desire is its role in causing actions, and another holds that the essence of desire is its role in causing feelings. These two types of theories identify desire with something that is familiar to everyone who has desires. The third type of theory holds that the essence of desire is its role in causing a certain form of unconscious learning. This type of theory strongly distinguishes what desires are, in themselves, from the very familiar effects that we all assume desires cause.

Action-Based Theories

Desires obviously play an important role in action. A sheep will walk from one place to another because it desires to eat grass, and a person will spend years in law school because she desires to fight environmental degradation. Action-based theories of desire make this role the essence of desire. One such theory (defended recently by philosopher Michael Smith) holds that to desire some state of affairs, \( p \), is to be in a state that makes it true that one is disposed to take action A if it seems that taking A will bring about \( p \) (or make \( p \) more likely). Thus, if Katie desires to learn quantum mechanics, then she must be disposed to take actions that she believes will make her learn quantum mechanics, and if Katie is disposed to take actions that she believes will make her learn quantum mechanics, then it must be true that Katie desires to learn quantum mechanics.

Action-based theories are the most prominent philosophical theories of desire; some would say that some version of this theory is an obvious truth, no more disputable than that a bachelor is an unmarried man. Among philosophers who embrace an action-based theory of desire, the main question is whether all actions are caused by desires (as the theory just proposed requires) or whether some actions might be caused by something other than a desire (such as a sense of moral obligation).

Feeling-Based Theories

Action-based theories of desire have been dominant from the 20th century until the present, but earlier philosophers often emphasized the connection between desires and feelings. Desires might be one cause of actions, but what is essential to a desire is the way it makes you feel, according to this sort of theorist. One such theory (defended recently by philosopher Galen Strawson) holds that to desire some state of affairs, \( p \), is to be in a state that makes it true that one is disposed to feel pleasure if it seems that \( p \) and to feel displeasure if it seems that not-\( p \). Thus, if Katie desires to learn quantum mechanics, then she
must be disposed to feel pleasure if she seems to be learning it, and if Katie is disposed to feel pleasure at the idea of learning quantum mechanics, then it must be true that Katie desires to learn it.

Proponents of feeling-based theories can make a distinction not available to proponents of action-based theories. The feeling-based theorist can hold that some people are (for example) kind because they desire to be kind. These (they can say) are the people who enjoy doing something kind. Other people are kind only because they feel they have a moral obligation to be so, but they do not desire to be kind at all. These (they can say) are the people who do not enjoy being kind but who do it anyway.

Learning-Based Theories

Philosopher Fred Dretske has initiated investigation into a third sort of theory of desire, which can be called a learning-based theory of desire. Learning-based theories of desire hold that apparent desire satisfaction causes a specific form of learning, known sometimes as reward learning or contingency-based learning. This form of learning causes an unconscious shift in an organism’s tendencies to go from one state of mind to another. Specifically, if one mental state (such as seeing a light while feeling hungry) causes another mental state (such as trying to push a lever) and then some reward learning takes place (caused, perhaps, by the arrival of some food after the lever was pressed), then the result of the reward learning will be to increase the power of the first mental state (seeing a light while feeling hungry) to cause the second mental state (trying to push the lever). Though this form of learning is best known from behaviorism, where it is called operant conditioning, there is nothing specifically behavioral about it, and there have been observations of reward learning modifying perceptual and cognitive associations as well as behavioral ones.

A basic learning-based theory of desire holds that to desire some state of affairs, $p$, is to be disposed to engage in unconscious reward learning if one believes that $p$. Thus, if Katie desires to learn quantum mechanics, then she must be disposed to go through reward-based learning when she believes she is learning it, and if she is disposed to go through reward-based learning when she believes she is learning quantum mechanics, then she must desire to learn it.

The obvious objection to a learning-based theory of desire is that learning, even the specific sort of learning in question, does not seem to have anything to do with desiring something. Defenders of learning-based theories have pointed out, however, that the brain’s reward-based learning system is the unique common cause of the actions, feelings, and thought processes associated with desiring. That is, although desires do not seem to have anything to do with reward-based learning, scientific investigation reveals that the brain’s reward-based learning system causes the effects that we normally think of as the effects of desires.

Neuroscientific Underpinnings

Science in general, and neuroscience in particular, has been reluctant to develop theories of desire. Perhaps this has been because there was not enough consensus about the nature of desire in, for example, philosophy in order for scientists to know what exactly they would be looking for (an exception has been the insightful work of Kent Berridge in neuropsychology). But in the years since 1980, neuroscience has developed extensive theories of action, feelings, and thought processes. It has also developed a sophisticated theory of reward learning. So a researcher who is interested in desire has a great deal to learn from neuroscience.

Neuroscience and Reward Learning

The brain contains a reward-learning system that releases dopamine. According to research undertaken by Wolfram Schultz and colleagues, this system releases dopamine at a baseline rate, which can then fluctuate on a short timescale so that dopamine release rises or drops relative to the baseline rate. The release of dopamine by the brain’s reward system expresses a simple calculation: actual reward minus expected reward (where expectations here appear to be unconscious rather than consciously calculated and where reward is shorthand for food for hungry organisms, water for thirsty organisms, and so on). Thus, a rat expecting nothing at time $t$, when food arrives, displays a rise in dopamine release, while a rat that fully expects food at time $t$, when food arrives, displays no change at that moment in dopamine release, and a rat that fully expects food at time $t$, when no food at all arrives, displays a drop in dopamine release.
The learning effects of dopamine have been documented both by direct observation of neural connections induced by dopamine released from the reward system (and connection formation blocked by dopamine antagonists) and by observation of the existence of distinct learning systems, some of which are dopamine dependent and some of which are not.

As a mediator of reward learning, the dopamine released by the brain’s reward system has incremental effects on dispositions to act, perceive, and cognize in certain ways. But the reward system has other effects also.

**Neuroscience and Pleasure**

One of the effects of the reward system is on pleasure. Dopamine release (or receipt) has been observed to be the means by which many pleasure-causing drugs cause pleasure: This is the mechanism of action for MDMA (ecstasy), amphetamines, and cocaine, for instance. Some (such as Carolyn Morillo) have suggested that the release of dopamine is the same thing as pleasure, but as Kent Berridge has argued, this cannot explain the full range of facts, such as the fact that certain drugs (opiates and benzodiazepines) appear to affect pleasure independently of dopamine and the fact that some experimental conditions that stimulate dopamine release and behavior appear to be unpleasant. The neural realization of states of pleasure (and displeasure) is probably causally downstream from dopamine release, and a number of neural sites (the nucleus accumbens, the ventral pallidum, the insular cortex, and the perigenual region of the anterior cingulate) have been suggested.

**Neuroscience and Action**

Another of the effects of the reward system is on action. Parkinson’s disease reveals how dependent human beings are on the dopamine released by the reward system for the capacity to act: Parkinson’s disease is caused by excessive death of dopamine-releasing cells in the reward system, and its effects include decreased ability to initiate action and (at the limit) complete paralysis of action.

These effects of the reward system are mediated by a key set of structures in action production known as the basal ganglia. The basal ganglia appear to coordinate action selection, combining information about what is happening in the world (information from perception and cognition) with information about reward and (perhaps, though it remains less well studied) punishment, with this latter information arriving from the reward system in the form of dopamine. At a high level of abstraction, it can be said that the motor systems of the basal ganglia take in output about how things stand and how rewarding they are, and it produces as output an action selected from those actions that are possible in context.

Interestingly, the power of the reward system to influence overt action strongly suggests a related influence over thought: Outputs from the basal ganglia appear to influence voluntary thought as well as voluntary bodily movement. This is also suggested by known cognitive impairments caused by Parkinson’s disease and by experiments linking rewards to modified thought processes.

**Integrating Philosophy and Neuroscience**

The neuroscientific facts just laid out strongly suggest that the brain’s reward system is involved in desire in some manner. The reward system is involved in causing action, it is involved in causing feelings, and it is involved in affecting thought processes. But what role should the reward system be given? All three philosophical theories of desire remain viable.

To support an action-based theory of desire, a theorist will claim that a desire for some conceivable state of affairs, p, exists when there is a connection between a neural representation of that state of affairs, through the reward system, into the action-production system such that, when the organism thinks some action would lead to p, the organism tends to take that action. This is the position defended by Kent Berridge after review of the role of the reward system in causing action (“wanting”) and feelings (“liking”). To support a feeling-based theory of desire, a theorist would claim that a desire for p exists when there is a connection between a neural representation of that state of affairs and pleasure, whether through the reward system (one possibility) or otherwise. And to support a learning-based theory of desire, a theorist would claim that a desire for p is whatever is the normal common cause of actions and feelings and also claim that the brain’s reward system is that common cause.

Although controversies about desire are far from resolved, it would appear that leading theories of desire link it, at least in part, to a common underlying...
neural structure. Which part of this structure is most important for desire itself depends on which theory is right, and that appears to remain largely a philosophical issue. But if philosophical theorizing can go forward constrained by scientific knowledge, there is some hope for progress.

Timothy Schroeder

See also Decision Making, Neural Underpinnings; Decision Theory, Philosophical Perspectives; Emotion, Structural Approaches; Happiness; Philosophy of Action; Reinforcement Learning, Psychological Perspectives

Further Readings


**Dictator Game**

This entry introduces and defines the Dictator Game (DG), a commonly used game in economic experiments. The DG was first used by Robert Forsythe, Joel L. Horowitz, N. E. Savin, and Martin Sefton in 1994, in which two anonymous players split money provided to only one of them. The person who receives money, the “dictator,” is given an opportunity to share this with the person who received nothing, the “recipient.” The DG was designed to reduce complex human behavior to simple steps to test game-theoretical predictions that assume rational utility maximization (i.e., maximization of the value of the money kept) based on mathematical calculations. It was assumed that dictators would not share money with anonymous strangers without some enforcement or punishment. It was a great surprise to learn that people shared, often splitting the money equally among themselves.

When two people who know each other or are related play the DG, it is expected that they will share. Standard economic theories, such as the neoclassical theory, for example, do not suggest that when playing with kin, one should maximize wealth at the expense of that kin. Even if the players don’t know each other, if the game is played repeatedly and if the player’s identity may be revealed later, economic theory suggests that there is reason to share.

![Dictator Game](image_url)
Building good reputation is a valuable asset, whose utility may be maximized. However, when complete strangers play a single game (one shot) with forever hidden identity, standard economic theories assume rational choices and expect the player to maximize wealth to self and to not share. However, in experiments played in various countries and cultures around the world, researchers found that the majority of people share with anonymous strangers in one-shot games. The game theoretic prediction that the “dictator keeps everything” happens only about 20% of the time. There is some cultural variation in how much money dictators keep. In some societies, zero offers simply do not exist. Dictators typically share between 20 and 50% of the money with anonymous recipients. The amount of money given to dictators as an endowment varies from a few dollars to substantial sums, sometimes several hundreds of dollars. Still, the size of the endowment seems to make little difference; 80% of the people share regardless how big an endowment they start with, and the percentage of money shared is stable.

Because in the DG sharing of the money is voluntary and the anonymous recipient benefits at the expense of the dictator, this game represents a classic example of altruistic behavior. There is gender bias in assessing who might behave more altruistically. When two boxes, one labeled from men and the other from women containing dictator donations, are placed in front of recipients, 80% of women and nearly 50% of men recipients prefer to choose from the women box. Replacing human recipients with charities reduces donations. This indicates that there is something special about sharing specifically with people even if they are anonymous and even if the opportunity to share surfaces only once.

**Dictator Game and the Mind**

Some of the factors that make the DG special surfaced when the game was modified. Rather than dictators giving to recipients, recipients were to take as much money as they wanted from the dictators’ endowment. If giving and taking can equal the act of sharing, giving and taking amounts would be identical. Recipients who take would take about as much as the dictators would have given to them. Mathematically speaking there is no difference between a person giving or another taking that same amount; all participants end up with the same money. However, when the recipients were given the option to take, they took more than they would have been given, often taking more than half of the endowment away from the dictator. It is apparent that adding and subtracting have strikingly different utilities associated with them in the human brain. While giving is altruistic, taking is selfish. The act of giving and taking activates different decision-making processes in the brain. Social consideration of sharing in the DG changes when the process of sharing changes.

Modern technologies, such as functional magnetic resonance imaging (fMRI) and other scanning processes, allow researchers to examine human neural processes while subjects play economic games. It is now understood that neural representations of various emotional states as well as genetic influences on those states guide human decision making. Receiving small amounts of money is seen as unfair by recipients and this activates the insula, a region of the brain that is associated with negative emotional states such as disgust, pain, hunger, and thirst. The mesolimbic reward center of the brain is a key player in decision making. It is activated the same way when making donations as when receiving money. In other words, giving money voluntarily may feel as good as receiving money. This explains why players give money to strangers but does not explain the differences in the amount of money given. Other brain regions, such as medial orbitofrontal–subgenual and lateral orbitofrontal areas, which play key roles in social attachment, are also activated in the donating process. This may show the social component and may help explain why, when people play with computer partners instead of human counterparts, they seem less upset by unfair offers—these regions of the brain may not be activated as intensely when not playing with a human partner. The need for social attachment is an important factor in the amount of money shared, which is reflected by the cultural influences over shared amounts.

Emotional states are also guided by hormonal influences. Neuropeptides (hormones in the brain) play key roles in social attachment. Arginine vasopressin (AVP) is a neuropeptide that regulates social attachment and kin protection. In a study by Ariel Knafo and colleagues, genetic examination of AVP receptors showed that people with long variants of AVPR1a RS3 repeats donate larger sums to
Discourse Processing, Models of

Everyday discourse experiences are rich and varied; they include perusals of newspaper headlines, gossip with friends about recent events, careful analyses of the results reported in research articles, establishing common ground with a new acquaintance, and relaxing moments immersed in a favorite novel. Success during these activities necessitates the recruitment of what readers, speakers, and listeners already know in the service of extrapolating meaning from linguistic constituents. The mental representations and cognitive mechanisms that underlie such language- and communication-based activities have been the focus of investigations in the field of discourse processing. These investigations have examined the moment-by-moment processes (e.g., prior knowledge activation) and final products (e.g., what has been learned) of linguistic experiences. An important element of this work has been the development of conceptual models that describe the nature of, and contributors to, discourse comprehension and production.

To exemplify this work, the focus here will be on the development and contributions of models in the domain of text processing. Contemporary investigations in this area often take as an inspirational starting point the seminal work of Walter Kintsch and Teun van Dijk and their tripartite model of text comprehension. According to the model, comprehenders can encode information they hear or read into memory in potentially three ways: (a) a surface-level representation, which stores the exact information presented in a text; (b) the textbase or propositional level, which includes the ideas conveyed by the surface form; and (c) the situation level, which goes beyond the text itself to include knowledge that readers might bring to bear on the texts they read. The power of the model comes from its utility in fostering predictions about the types of mental representations and concomitant understandings generated as a function of, as examples, readers’ goals, the difficulty of the text, or the familiarity of the text topic. Applications of the model have made important contributions to the study of reader memory, the types of inferences constructed during reading, and the rhetorical and structural features that can enhance learning from text.

All such model applications have held, as issues of core importance, the need to understand how prior knowledge and unfolding discourse separately contribute to and mutually interact during linguistic experiences. Based on accumulated empirical evidence, current models (including Kintsch’s construction-integration model and recent iterations of memory-based and constructionist views) have converged on a two-stage process of knowledge activation and integration. In the first stage, incoming stimuli broadly and automatically activate concepts in memory. This initial activation of concepts is unconstrained, which means that both relevant and irrelevant information can be primed by linguistic input. In the second stage, relevant concepts are selected from the unconstrained group, and the chosen elements are integrated into the unfolding mental representation of the discourse. Similar versions of this model have been articulated for lexical access, speech planning and production, literal and nonliteral language comprehension, and conversational grounding. A crucial consideration to emerge from this work is the notion that discourse processing is dynamic; concepts in memory wax and wane in activation over the course of, for example, a text...
experience, and these activations influence the information that will be integrated into memory for use after the experience has concluded.

The models described thus far have received considerable empirical support, through methodologies including think-aloud protocols, tests of concept accessibility (e.g., lexical decision and naming), profiles of cortical activity (e.g., event-related potentials and hemispheric lateralization studies), recall tasks, and measures of reading latencies. An additional test has involved building computational simulations of core model mechanisms and comparing simulation performance to human data. One example is the landscape model developed by Paul van den Broek and his colleagues, which simulates the activation patterns of concepts over the two-stage process described above. The model's architecture incorporates processing contributions from current text input, the immediately preceding text input, an episodic representation of all of the previous text information, and the knowledge base of a hypothetical reader. These sources are used to calculate the dynamic activation levels of text concepts both during and after reading. The degree to which the calculated output aligns with the data of actual readers provides one type of existence proof for the core mechanisms and assumptions of the model. Simulations can also provide the opportunity to test novel hypotheses. For example, model architectures can be changed to vary the reading skill or prior knowledge available to the system by constraining the contributions of input sources. These kinds of modifications provide an analysis of how the structural components of a model can contribute to discourse processing. To date, several computational models have proved impressive in their simulations of reader performance.

Models of discourse processing, both conceptual and computational, are also used to consider the loci of a variety of processing difficulties. For example, students in school settings often struggle to comprehend classroom discussions and materials. Several models have been implemented in consideration of whether difficulties are, at least partially, due to failures of particular discourse processing mechanisms. Models have also been utilized to suggest approaches for the remediation of problems through structured interventions or the design of learning tools (such as Arthur Graesser's AutoTutor). The continued development and refinement of these models will be beneficial for future attempts at describing, explaining, and enhancing the processes that underlie discourse comprehension and production.

David N. Rapp

See also Conversation and Dialogue; Eye Movements During Reading; Inferences in Language Comprehension; Working Memory in Language Processing

Further Readings


DIFFERENTIAL LEARNING, TRAINING METHODS

The term discrimination learning—learning to respond differentially to different stimuli—encompasses a diverse range of different types of learning, such as perceptual learning, concept learning, and language learning. The practical goal of discrimination training often involves complex discriminations, such as when pilots learn to recognize potentially hazardous patterns on complicated instrument panels. The greatest challenge in discrimination training—aside from the initial problem of perceiving differences between categories—is to accomplish long-term learning that transfers across situations and stimuli. Training techniques that produce good performance can be misleading; difficult training is sometimes more effective than easier training in the long term.

Discrimination learning occurs at many levels, such as perceptual learning, higher level concept and category learning, and language learning. A baby
experiences all three types of learning: She has to learn basic perceptual categories (e.g., discriminating between *ba* vs. *ga*), higher level concepts (e.g., friends and relatives vs. strangers) and linguistic labels for these categories and concepts (e.g., *dog* refers to furry pets but not cats).

Learning to discriminate often involves a process of differentiation and unitization. Differentiation involves learning to separate similar stimuli into different categories (e.g., ravens vs. crows); unitization involves learning to group stimuli into larger units and to discriminate based on the unit (e.g., seeing a whole face instead of separate facial features). Unitization and differentiation may appear to be opposite processes, but both involve learning to perceive information at the appropriate level of representation.

**Real-World Applications**

In real-world applications, the usual goal of discrimination training is to teach people or animals complex discriminations. For example, drug-sniffing dogs learn to discriminate among complex odor compounds in order to respond appropriately. Athletes, such as quarterbacks in American football, goalkeepers in hockey, and batters in baseball, need to be able to discriminate very complex and fast-moving patterns of stimuli and respond appropriately. In athletics and more generally, perceptual learning works in concert with motor learning, which requires discriminating among internal sensations, such as the proprioceptive feeling of an ace tennis serve versus a serve that goes into the net.

In school, learning occurs primarily via direct instruction: Books and teachers provide information, and students try to understand and remember that information. Direct instruction alone is not sufficient to promote many types of discrimination learning, however, which often requires a combination of instruction and experience. Verbal descriptions of a genre of music, for example, may be of little help compared with actual experience listening to the music.

**Factors Affecting Discrimination Learning**

In most cases, the goal of discrimination training is long-term transfer—that is, teaching learners in a way that will be relevant when they encounter novel stimuli and new situations. For example, a soldier taught to distinguish enemy soldiers from civilians during basic training needs to be able to transfer that learning to the battlefield. Training techniques that make training conditions difficult—and thereby decrease performance levels during training—often increase long-term learning and transfer. Such techniques are referred to as desirable difficulties. In learning to recognize patterns, for example, increasing the variability between members of a category during training decreases training performance but positively affects transfer.

Trainers and trainees often mistake short-term performance during training as evidence of long-term learning and therefore perceive difficulty during training as an impediment when it is actually an advantage. For example, people think they learn artists’ styles best when the artists are presented for study one at a time, whereas in reality mixing various artists’ paintings together enhances categorization of novel paintings. It is important not to be fooled by training techniques that enhance short-term training performance at the expense of long-term learning and transfer.

Although difficult training can be vital, especially if the goal is to train difficult discriminations, too much difficulty can prevent learners from identifying categories’ essential features. For that reason, perceptual training often begins with easy discriminations that adaptively increase in difficulty during training. However, discrimination learning can also occur when discrimination is impossibly difficult; asking people to discriminate between physically identical tones, for example, enhances transfer to a discrimination task on which tones vary.

At relatively early stages of perceptual processing, such as learning to make fine distinctions between lines at various orientations, transfer is frequently difficult to accomplish. Learning does not necessarily even transfer when an old task is presented in a new spatial or retinal location. However, long-term training can increase basic sensitivity in a way that can be transferred; for example, violinists show more activity in the sensory cortex when touched on their left hand (i.e., the violin neck hand) than when touched on their right hand (i.e., the bow hand).

Some forms of discrimination learning seem to benefit from a period of sleep following training, which appears to aid memory consolidation. In a simple visual discrimination task, for example,
people who are allowed to sleep normally show improvement 3 days after initial training. People who are not allowed to sleep for 30 hours following training and are then tested after two nights of sleep show no improvement compared to baseline performance.

Nate Kornell

See also Categorization, Psychological Perspectives; Category Learning, Computational Perspectives; Concepts, Development of; Desirable Difficulties Perspective on Learning; Human Classification Learning

Further Readings


Disfluencies: Comprehension Processes

Human speech is far from perfect: Approximately one in six words of spontaneous speech is affected by false starts, repetitions, or hesitations. A traditional view is that these disfluencies are ignored by listeners, and it is true that listeners are not good at identifying occurrences of disfluency in recorded speech. However, evidence suggests that disfluencies, particularly fillers such as *uh* or *um*, have demonstrable effects on the comprehension of speech. This entry summarizes those effects, which can be broadly broken down into three classes. First, disfluencies have been shown to affect listeners’ judgments about speakers’ confidence in what they are saying (“feeling of another’s knowing”). Second, a number of studies have suggested that the processes of prediction and attention that underlie comprehension can be influenced by disfluency. Third, disfluencies affect the eventual representation of what was said. Listeners may not process them consciously, but disfluencies are integral to spoken communication. The entry concludes with a brief consideration of whether different disfluencies give rise to different effects.

Judgments About the Speaker

When speakers are unsure of an answer to a question they tend to hesitate, and these hesitations may be marked by utterance-initial disfluency. Speakers are asked to rate their own confidence in their answers (their *feeling of knowing* or FOK), produce ratings that correlate negatively with disfluency. Listeners are equally able to make this kind of judgment: Their ratings of FOAK (*feeling of another’s knowing*) are lower when answers are preceded by silence and lower still when the silence is replaced by a filler-silence combination of the same length. These findings provide evidence that disfluency can influence the judgments that participants are asked to make. However, they do not directly answer the question of whether disfluency routinely affects language comprehension.

Effects on the Comprehension Process

Listeners are faster to identify target words following *uh* in recorded speech than when the *uh* are excised and are faster to respond to instructions to press a given button when those instructions are repaired (*yel-uh-orange*). Disfluencies appear to have a helpful effect on the ongoing comprehension of speech, at least in carrying out specific tasks. More compelling evidence comes from studies in which listeners’ eye movements are recorded as they respond to instructions to manipulate objects on a computer display. In line with other work on language comprehension, listeners appear to predict what is most likely to be mentioned at any given stage, and these predictions are reflected in anticipatory eye movements to the objects depicted. When the instructions are disfluent, participants tend to fixate earlier on objects that are hard for the speaker to describe or name (such as visually complex abstract objects). When listeners are told that the speaker is aphasic and prone to random disfluency, the effects disappear, suggesting that they can be directly attributed to the listeners’ theories about what (normal) speakers find difficult to say.
Additional work has used event-related potentials (ERPs) to examine whether disfluencies routinely affect how listeners process speech. In these studies, there is no specific task to carry out or set of objects referred to. Participants simply listen to recorded utterances, which include some target words that are surprising (difficult to integrate semantically or acoustically deviant in context). Relative to unsurprising control words the targets give rise to clear ERP effects (an N400 for semantic and a P300 for acoustic surprise), but these effects are attenuated when the target words are immediately preceded by an *uh* disfluency. These ERP findings implicate prediction (not being able to predict what comes next may reduce the semantic surprise and hence the N400) and attention (the P300 reflects attentional orientation to input; because attention is oriented by the disfluency, *uh* results in a P300 and the subsequent P300 is diminished). In other words, disfluency indicates that the words that follow may not be predictable, which causes listeners to attend more closely to what follows.

**Effects on Representation**

When tested subsequent to the comprehension phases of the ERP experiments outlined above, participants were more likely to recognize having heard target words when those words had been preceded by an *uh*, consistent with the view that attention would have been heightened in these circumstances. As well as affecting language comprehension as the process unfolds, disfluencies have consequences for its outcome. Further evidence that disfluencies can affect the representation of what is said comes from experiments in which participants listened to sentences that sometimes included disfluencies (*uh uh*) at potential clause boundaries. In these experiments, participants were asked to decide whether the sentences they heard were grammatical or not. They were less likely to rate the sentences as grammatical in cases where the disfluency occurred at a boundary that turned out to be inconsistent with the correct interpretation of the sentence, showing a direct influence of disfluency on interpretation.

**All Disfluencies Are Not Equal**

Perhaps surprisingly, the grammaticality judgment effect is also found when sounds that are not disfluencies (such as dogs barking) are used at potential clause boundaries. However, other studies have repeatedly led to the conclusion that all disfluencies may not be equal. For example, the benefit of *uh* in target word identification does not appear to extend to *um*, and the ERP effects associated with *uh* are not completely replicated when the disfluency is a repetition. At the time of writing, the question of whether different types of disfluency affect listeners differently is unresolved.

*Martin Corley*

See also Conversation and Dialogue; Planning in Language Production; Prosody in Production; Speech Perception

**Further Readings**


**Disgust**

Most contemporary theorists assume that emotions have evolved because of the adaptive functions they provide to the individual experiencing them. Two of the key functions of emotions are first, to facilitate rapid responses toward potentially dangerous aspects of the environment, and second, to facilitate social interactions with other individuals. Disgust plays an important role in both respects: Although it was originally concerned with the avoidance of potentially harmful substances, it is also involved in the social enforcement of norms.
Disgust as a Basic Emotion

Theoretical debates continue regarding the existence of a fixed number of basic emotions; however, all proposed lists of basic emotions include disgust as one of them, usually together with happiness, sadness, anger, fear, and surprise. Disgust is considered to have evolved in the context of food consumption as an emotion that protects from various contaminants, such as spoiled food, bad taste, and unpleasant odors, in order to prevent ingestion of substances that may potentially be harmful. As is the case for other emotions, disgust involves a characteristic subjective experience, a facial expression, and some action tendencies. Because it is concerned with appetitive responses related to food consumption, a disgust response consists of literally moving offensive substances away from the self and can include expulsion of bad substances by spitting them out, or, if ingestion has already occurred, vomiting. The disgust facial expression involves the mouth gaping open, which facilitates food rejection and shutting down sensory intake of offensive smells by pulling up the nose and raising the upper lip. A subjective feeling of revulsion, and possibly nausea, accompanies this rejection response. In addition to the avoidance of toxins that might be ingested by mouth, disgust also guards against the touching of contaminated substances and of parasites and pathogens that might be harmful by contact, thus avoiding potential infection and disease.

Many emotions can be modified through cognitive reappraisal; that is, the subjective experience can be altered by cognitive strategies, such as giving a different interpretation to an emotion-eliciting stimulus. In contrast, feelings of disgust are relatively immune to cognitive reappraisal, with people finding it difficult to perceive a disgusting stimulus in a non-disgusting way. For example, because of preexisting associations with disgusting objects, most people will refuse to drink apple juice from a brand-new bedpan or eat chocolate fudge in the shape of dog feces. Further, people tend to engage in magical thinking with respect to disgusting objects and situations, such that perceived contamination is treated as if it were contagious, even if this is not objectively the case. In addition, objects are often considered to have some essence of contamination such that once something is disgusting, it remains so even after intensive cleaning efforts. A certain asymmetry exists such that any object can easily go from being clean to being contaminated, but the reverse is not true, because to some extent contamination is considered irreversible.

Types of Disgust

Three distinct types of disgust have been identified: first, core disgust; second, animal nature disgust; and third, sociomoral disgust. First, the most basic type of disgust, core disgust, describes the function of disgust that is linked to food and consumption, as described above. In addition, animal nature disgust describes people’s emotional response of disgust and repulsion when reminded of the fact that human beings are phylogenetically similar to nonhuman animals. Consequently, physiological processes related to bodily functions, especially those related to bodily wastes and reproduction, are considered animal-like and therefore disgusting. A strong desire exists to view human beings as enjoying a special status among other creatures, and people go to great efforts to distinguish themselves from animals by implementation of social and cultural norms. When the boundaries between human beings and animals are blurred, people feel debased. Presumably, the main reason for the discomfort associated with reminders of human beings’ animal nature is that they make salient the fact that, just like any other creature, every human being is mortal and will eventually die.

As is common in evolutionary history, structures that initially developed for one use were co-opted for other uses. Sociomoral disgust is based on basic physical disgust that was extended to more abstract contexts such that people find situations disgusting in which moral standards are violated. Thus, disgust is a reaction to offensive objects, as well as offensive actions. Supporting the close link between the different types of disgust, neuroimaging studies have shown that the same brain structures are implicated in the experience of core disgust and sociomoral disgust. Further, sometimes people can mistake physical disgust for moral disgust: In experimental manipulations, research participants were induced to feel physical disgust, for example, by being exposed to a bad smell. Relative to participants in neutral control conditions, participants experiencing disgust gave more severe judgments on a variety of moral transgressions, presumably because they conflated experiences of physical and moral disgust. Thus, disgust is an emotion that acts as guardian of both the body and the soul, ensuring that no physical or moral contamination taints the self.
Cultural and Individual Differences

Across many cultures, the same verbal and nonverbal expressions used to reject physically disgusting things are also used to reject certain kinds of socially inappropriate people and behaviors. However, in addition to substances and behaviors that are universally rejected, norms of what is considered disgusting vary by culture. For example, consumption of meat of some animals (e.g., cows, pigs, dogs) is considered acceptable in some cultures but unacceptable in others. In addition, cultural norms regarding aspects of hygiene (e.g., toilet training) shape children’s disgust responses.

Although the processes related to disgust that are described above appear relatively universal, some people are more susceptible to the experience of disgust than others. Individuals high on so-called disgust sensitivity tend to show a higher prevalence of anxiety disorders such as phobias and obsessive-compulsive disorders. Thus, disgust is an example of an emotional response that is normally highly adaptive, but in psychopathology it can become extreme and thus maladaptive.

Simone Schnall

See also Emotion, Cultural Perspectives; Emotion, Structural Approaches; Emotion and Moral Judgment; Facial Expressions, Emotional

Further Readings


Disjunctive Theory of Perception

The main claim of the philosophical view of perception commonly labeled as the “disjunctivist theory” is that veridical perception (perceiving things correctly) and hallucination are very different, even when they are phenomenally indistinguishable. A perceptual experience is either a veridical experience or a hallucination and, importantly, there is no common denominator between these two disjuncts. In order to understand what disjunctivism entails, it needs to be examined not only how these claims are to be cashed out but also what motivates them.

Perception: Representation or Relation?

There are two very different ways of thinking about perception. The first one is this. Perceptual experiences are representations: They represent the world as being a certain way. They have content, which may or may not be different from the content of beliefs. They represent objects as having properties, sometimes veridically, sometimes not.

According to the other influential view, perception is a relation between the agent and the perceived object. Perceived objects are literally constituents of our perceptual experiences and not of the contents thereof. Perceptual experiences are not representations. Following John Campbell, we can label these views the representational and the relational view, respectively.

One of the explanatory advantages of describing perceptual experiences as representations is that it allows us to treat veridical and nonveridical cases of perception in similar manner; thus, this could be termed the common factor view. Our beliefs can be incorrect and so can our perceptual experiences. The representational view can give a simple explanation for this: Both beliefs and perceptual experiences can fail to represent correctly; both can misrepresent. I may hallucinate that there is an apple on my desk. In this case, I have a perceptual experience that misrepresents. It represents an apple in front of me but in fact there is no apple in front of me. If we accept the representational view, hallucinations and illusions are considered to be perceptual experiences that misrepresent their objects.

Hence, the representationalist has a neat story to tell when it comes to the relation between veridical and nonveridical experiences. But what can the relationalist say? As the main claim of the relational view is that the perceived token object is a constituent of our veridical perceptual experiences, if we are hallucinating, there is no perceived token object that could be this constituent of our experience: Veridical experience is a relation between the perceiver and the perceived token object. But
hallucination cannot be conceived of as a relation, as one of the two \textit{relata} (the perceived object) is missing. In other words, veridical experience and hallucination are very different (but maybe indistinguishable) experiences.

This relationalist way of accounting for the relation between veridical and nonveridical experiences is called \textit{disjunctivism}.

\textbf{What Disjunctivism Entails}

We have to be careful about what is meant by the identity or difference of experiences in order to properly interpret the main claim of disjunctivism. What the debate between the disjunctivist and the common factor view is about is whether my veridical experience of a pillow and my indistinguishable hallucination of a pillow are of the same type. But then this disagreement no longer seems very clear, as there are many ways of categorizing experiences as belonging to different types. Even the disjunctivists would agree that we can do so in such a way that the two token experiences would both belong to the same type, say, the type of experiences in general. And even the common factor theorist could say that there are ways of sorting these two experiences into very narrowly defined types so that they end up belonging to different types.

It has been suggested by M. G. F. Martin that the real question is whether these two experiences belong not just to the same type but whether they belong to “the same fundamental kind.” The representational view says they do; the relational view says they don’t. Belonging to a “fundamental kind” is supposed to “tell what essentially the event or episode is.” Those who are suspicious of anything fundamental or essential will not find these considerations too compelling.

\textbf{Varieties of Disjunctivism}

A further complication is that there are other kinds of perceptual experiences besides veridical and hallucinatory ones. Importantly, there are illusory perceptual experiences, where different properties are attributed to the perceived object than what they in fact have. The question is, Which side of the disjunction do illusory perceptual experiences fall? Depending on how we answer this question, we get different versions of disjunctivism: the \textit{VI versus H} and the \textit{V versus IH} versions (where \textit{V} stands for veridical perception, \textit{I} for illusion, and \textit{H} for hallucination).

Yet another complication is that disjunctivism, as stated here, is a purely negative account of hallucination. It states that hallucination is nothing like veridical perception, but it does not give a positive account of how we should think about hallucination. At this point, there is another split in the disjunctivist camp: The negative disjunctivists insist that there is nothing more to say about hallucination, whereas the positive disjunctivists aim to come up with a positive account of what constitutes hallucination—and this account needs to be able to explain the experiential similarity between the veridical and the hallucinatory cases.

Finally, it is important to note that disjunctivism can be combined with representationalism about the nonveridical cases. Although disjunctivism was originally a supplement to the relational view in order to explain nonveridical perception without appealing to misrepresentation, the view can be detached from its relationalist origins. One such possible account would be to claim that veridical experiences are genuine relations between the perceiver and the perceived object; they do not represent anything, whereas hallucinatory experiences are representational states with content. This view is sometimes referred to as \textit{disjunctive representationalism}.

\textit{Bence Nanay}

\textbf{See also} Conscious Thinking; Theory of Appearing

\textbf{Further Readings}


Dissent, Effects on Group Decisions

Dissent in groups means that different initial opinions or decision preferences are held before a particular issue is discussed. Although dissent is often experienced as negative, it has positive effects on group decision quality. In this entry, these positive effects are summarized and the psychological processes leading to them outlined; the downsides of dissent are briefly addressed; and some suggestions are offered about how dissent can be mimicked if authentic dissent is absent.

The Benefits of Dissent for Group Decision Making

Most people do not like dissent because dissent implies conflict, and conflict is undesirable. The expression of dissent lowers group members’ satisfaction with the process and the outcome of group decision making. Generally speaking, people value consensus and harmony in a group.

However, the absence of dissent can have disastrous consequences. The American psychologist Irving L. Janis analyzed famous decision fiascoes in U.S. foreign policy (such as the Bay of Pigs invasion in Cuba) using historical documents, biographies, and other archival data, and demonstrated that the groups that had made these disastrous decisions suffered from extreme consensus seeking and a lack of conflicting opinions in the group, leading to closed-mindedness, a severe underestimation of risks, and a faulty decision-making process. Janis contrasted these fiascoes with two successful decisions (e.g., the reaction to the Cuban Missile Crisis) and showed how in these cases procedures aimed at establishing robust debates in the group helped achieve a sound decision.

Interpretations of historical data, however, are likely to be biased by the hypotheses the researcher has in mind. Therefore, the question is what controlled experiments and field studies tell us about effects of dissent on group decision quality. Interestingly, this literature also indicates that dissent can be beneficial for group judgment and decision quality. Specifically, greater variety in individual judgments and decision preferences prior to group discussion is associated with better final group judgments and group decisions. This is particularly true for situations where the superiority of certain alternatives cannot be detected on the basis of the group members’ individual prediscussion information but requires the exchange and integration of members’ unique information during group discussion (so-called hidden profiles). In such situations, dissent fosters the detection of superior solutions that have not been in the focus of the group members’ attention before; these findings are supplemented by field studies showing that dissent is associated with higher innovativeness in teams.

Productive Dissent: How Does It Work?

The positive influence of dissent on group decision quality relies on two different effects. The first one is the proponent effect. It means that the more variety in prediscussion preferences is given, the more likely it is that the group will—by chance—contain members that already prefer the best or a nearby optimal solution prior to discussion, and the influence exerted by these members will benefit group decision quality. The second effect is the pure dissent effect. It means that even if all group members prefer suboptimal alternatives at the beginning, their disagreement makes it more likely that they will find the optimal solution in the group. Thus, a diversity of wrong opinions can sometimes generate the right idea.

Two different social-psychological processes mediate this pure dissent effect. On the one hand, dissent heightens discussion intensity: In groups with prediscussion dissent, more information is exchanged, information is repeated more often, and discussion lasts longer than in groups without prediscussion dissent. On the other hand, dissent lowers discussion bias: Groups with prediscussion dissent are more open to unique information that is new to most group members and focus less on information that supports the speakers’ initial preferences than groups without prediscussion dissent.

Dissent also has its downsides. As mentioned, it prolongs the decision process in groups, thereby reducing decision speed. Furthermore, dissent lowers group members’ satisfaction with the process and outcome of group decision making, which can reduce members’ willingness to implement the
solution finally found in the group and to work together in the same group again.

**Productive Dissent: How Can We Make It Work?**

Dissent effects can be facilitated. Differences in initial opinions only benefit the process and outcome of group decision making if they are expressed and, at least to a certain extent, upheld. Given the multiple barriers to the expression of dissent that can often be found in real groups such as, for example, directive leadership or norms that value harmony (instead of a critical debate), it cannot be taken for granted that dissenting opinions will be voiced. Therefore, high levels of participation in teams, meaning that group members feel free to and have the opportunity to contribute their own perspectives, have been shown to facilitate effects of dissent on team innovativeness.

It is not always possible to compose a group in a way that prediscussion dissent will be a given. Therefore, so-called dialectical decision techniques have been developed that aim at mimicking the effects of (authentic) dissent by means of installing a robust debate independent of the members’ real opinions. An example is the devil’s advocacy procedure, where a group member or a subgroup is assigned the role of the devil’s advocate who has to systematically criticize the proposals that are made by the group. As with authentic dissent, such contrived dissent has also been shown to be beneficial for the quality of group judgments and group decisions. However, in comparison to authentic dissent, both the benefits and the downsides of contrived dissent seem to be somewhat less pronounced.

_S Stefan Schulz-Hardt and Andreas Mojzisch_

*See also* Debiasing; Decision Improvement Technologies; Group Decision Making; Wisdom of Crowds Effect

Further Readings


**DISTRIBUTED COGNITION**

The term *distributed cognition* has been employed in a variety of theoretical perspectives, both in a weaker sense as a metaphor for coordinated social activity, as well as in more theoretically substantial forms in which the elements of the extended system act as a physical architecture for cognition. This entry will cover distributed cognition’s historical development within the social and cognitive sciences, orientation, areas of application, theoretical foundations, and methods of investigation.

**Origins and Orientation**

In its most fully fleshed out form, distributed cognition (or DCog) was developed primarily in University of California San Diego’s Department of Cognitive Science, by Edwin Hutchins at the Distributed Cognition and Human-Computer Interaction Laboratory. Its development alongside human-computer interaction is an important one, and it has been extensively used and promoted by this wider community, most notably by the psychologist and popular writer, Donald Norman. Analyses applying distributed cognition have been carried out on a variety of systems, including airline cockpits, air traffic control, ship navigation, computer programming, medical informatics, construction, and trawl fishing. Distributed cognition is closely related to activity theory (from which it has developed some of its terminology), situated action, and situated cognition. All of these emphasize the importance of the (external) context of the activity to varying degrees.

DCog expands the focus of cognitive activity away from the individual acting alone and unsupported, toward a system of people and tools as a unit of analysis. In its original conception, it was developed as a corrective to the forms of cognitive science prevalent at that time, which focused
entirely on the individual mind working in isolation. In opposition to this, DCog attempts to show how cognition is both a social and a cultural process and that traditional cognitive science may be attributing intelligence to internal mental structures when this is not the case. DCog may occur through enlisting external tools (e.g., through the use of a calculator, pen and paper, or maps) or by enlisting other actors in the problem-solving task. DCog also offers analytic and practical potential, showing how a system’s organization and other resources and constraints are assembled to produce intelligent action; these insights can be utilized in redesigning the structure and materials of these activity systems.

**Probing the Sociocultural Components of Cognition**

DCog retains the emphasis on information processing and problem solving developed from cognitive science, yet extends this approach to larger and distributed systems of activity that are situated in the real world and sensed and enacted through physical mechanisms. One concern dominates this, and draws from a largely neglected strand in cognitive science—culture—connecting the understandings, roles, and relationships between the human elements in the activity, or functional, system, and how it is shaped by the history of interaction between elements (human and nonhuman) within the system, drawing strongly from the cultural-historical tradition of activity theory. In the form of activity theory developed at the Moscow Institute of Psychology by Lev Vygotsky and Alexander Luria and refined by Alexei Leont’ev, many psychological processes are best understood as culturally and historically embedded by dint of being activity processes structured through communities of practice rather than as individual decisions determined independently by those involved. This perspective is developed further in DCog, which claims that most environments have what Hutchins calls a “cognitive ecology”; that is, they are sociocultural constructions, not arbitrary natural settings but artificially structured to support DCog processes.

This anthropological perspective draws physical artifacts, cognitive tools, and other people into the analytic frame and allows them to be examined as component features of the broadened cognitive system. DCog systems can involve multiple, nested functional systems operating at a lower level, rather like Russian dolls, each of which can be understood as goal-directed systems, interacting sequentially or in parallel with each other to form higher level goal-directed systems (albeit with different goals to the lower level systems). Naturally, studying distributed cognitive systems can require different methods of investigation to more traditional forms of study, and, rather than laboratory-based experiments, a qualitative, observational approach to uncovering naturalistic patterns of behavior, known as ethnography, has been extensively employed. Drawing from a rich tradition in the social sciences, ethnography was developed over the first half of the 20th century as the dominant form of research in cultural and social anthropology with canonical works by Bronislaw Malinowski, Margaret Mead, and, more recently, Clifford Geertz. Within DCog, the focus of such (cognitive) ethnographies lies in uncovering the organization of distributed cognitive systems, emphasizing the role of representational artifacts (i.e., external to the mind) and the ways that these representations are modified through their use and sequences of action.

**A Distributed Cognitive Architecture**

Within the framework of distributed cognition, cognition is seen as a computation, with inputs, outputs, and a process of mapping information through a variety of coordination mechanisms across representational media. Importantly, knowledge structures may be external to the mind. In taking the position of a DCog analyst, involving new representational artifacts or making changes to the practical organization of work does not simply augment the system, but actively changes the nature of the computation. Seen from such a systems approach, tools do not make people better at tasks; rather they change the task to one that they are better able to achieve.

DCog implies that human cognition may differ from the conceptions that psychologists have focused on in the past and have attempted to examine using laboratory experiments. By reassessing the critical role of external resources in cognition, DCog theorists have come to consider that a core role of mental activity lies in coordinating the various representations (which may be internal or external) that can be brought into alignment with one another to perform some sort of (broadly)
Divided Attention and Memory

goal-directed activity; that is, the role of mind is to shape the extended problem-solving system rather than to simply act as a “machine” to perform the computation itself.

The division of labor is an important feature of distributed cognition, which entails that tasks are broken down and allocated to specialized entities, and subsequently reincorporated together. In contrast to social and organizational theory, at an analytic level, DCog does not differentiate between the division of (cognitive) labor between human and nonhuman elements. Thus, within the DCog framework, functional systems may be composed of people and physical artifacts, and of social interactions between people that may, or may not, involve physical artifacts. In this sense, DCog offers a single, unified framework for the analysis of complex sociotechnical systems that does not require a range of analytic or theoretical techniques to be employed. For systems that are broadly social in nature, such complex systems tend to be described as socially distributed cognitive systems.

Within socially distributed cognitive systems, elements of the cognitive system lie not just in the knowledge and skills of individual agents, but also in the organization of those individuals, through the configuration of the tools that they use and their environment. Thus, cognition can be and is distributed socially, spatially, materially, and over time. At the same time, actors do not passively adapt to existing structures; they proactively structure these environments over time by organizing and reorganizing the physical and cognitive artifacts in the environment, and generating and transforming the social context.

Mark Perry

See also Collective Action; Extended Mind; Group Decision Making

Further Readings


Divided Attention and Memory

The dual-task or divided-attention (DA) technique can be used to aid our understanding of cognitive functioning by helping us to infer the type of resources and component processes required for a particular task. The logic in dual-task studies is that by comparing conditions in which attention is divided between two tasks, one can infer by the disruption in performance, relative to a nondistracting condition, whether the concurrent tasks require the same processing resources or system for representing information. This entry highlights studies that have used this technique to gain insight into several cognitive abilities.

Divided Attention Tasks in Perception and Working Memory Research

The dual-task technique has been used extensively to study the general processing capacity of humans. Daniel Kahneman’s view of human information processing, in which attention is described as drawing on a single, limited, pool of resources, was challenged based on dual-task research. While some cognitive tasks were shown to be difficult to combine, because they drained a single pool of attentional resources, others could be combined relatively easily, suggesting that multiple pools of attention exist. For example, repeating aloud (shadowing) sentences in a dichotic listening test is hampered when attention is divided with another task involving comprehension of simultaneously presented words. In contrast, Alan Allport and his colleagues found little interference when attention was divided between word shadowing and a task such as playing piano music or recognizing pictures. Along the same lines, Lee Brooks showed that reading, a visual task, was more difficult when performed concurrently with another task that required internal visualization (imagery), than it was when combined with one without a visualization component. Such results were incompatible with a single-resource view and led researchers
to propose that humans have multichannel processors or pools of attention resources; it is only when the same ones are required simultaneously for two tasks that performance will decline.

Marcel Kinsbourne and Robert Hicks attempted to understand the results of dual-task experiments by proposing that the degree of interference from simultaneous tasks is an inverse function of the functional distance between cerebral control centers. By contrast, research by Hal Pashler and his colleagues showed that competition for a central response channel forms a bottleneck that can account for much of the slowing that occurs when two tasks, be they memory, perception, or motor tasks, are combined.

The dual-task technique has also been used to elucidate the characteristics of working memory, which provides temporary storage and allows the active manipulation of information necessary for complex cognitive tasks such as arithmetic and reading comprehension. The model of working memory first proposed by Alan Baddeley and Graham Hitch in 1974 was influenced by various DA experiments and drew its support from them. The model posits a limited-capacity central executive component, which coordinates operations of two subsidiary storage systems, the phonological loop and visuospatial sketchpad. These subsystems provide temporary representation for verbal and visuospatial material, respectively. Support for the model came from Robert Logie’s studies showing a greater disruptive effect of unattended speech on short-term memory for words or digits than on memory for visuospatial information such as patterns on a visually presented grid, and vice versa if the task involved visual imagery and the interference was spatial.

**Divided Attention Tasks in Long-Term Memory Research**

The DA technique has also been used to characterize how humans encode, and later remember, information from the distant past, termed long-term memory. It is reasonably intuitive that our long-term memories will be disrupted easily when attention is divided between two tasks during learning—the encoding (study) phase. For example, listening to a conversation, watching images on a TV, or carrying out numerical calculations while studying a list of words for a memory test all hurt later retrieval of the target word list. DA at encoding has been shown consistently to produce substantial decrements in long-term memory performance, regardless of the type of distracting task used to divide attention. During initial encoding of information, reductions in general processing resources needed for conscious awareness, elaboration, and organization of information have a consistent negative effect on long-term memory.

More counterintuitive is the finding that, in the majority of research studies, DA during the retrieval (test) phase has much less of an impact. This is unexpected since most people allege that retrieving information from memory—be it the name of a movie, familiar face, or answer to an exam question—is an effortful task often thwarted by distraction. Whether DA at retrieval affects long-term memory may depend on two related factors: the resource demands of the particular memory test and the type of task that is performed concurrently. As at encoding, effects of DA at retrieval are large, regardless of type of concurrent task, but only when the memory test is one that benefits from executive type or strategic processing. Examples of these memory tests include recall of categorized word lists in which sorting words into groups, and later recalling the semantic categories, can boost recall of the word list; recall that involves discriminating the source or context associated with the target memories; and memory tests that require overcoming interference from other, related, items. Such large general effects of DA at retrieval are interpreted as resulting from competition between the distracting task and the memory test for general cognitive resources, presumably mediated by the frontal lobes. By contrast, other memory tests such as free (random order) recall, and cued recall in which prompts are given to guide retrieval, rely less on executive strategic processes mediated by the frontal lobes and more on direct associative ones mediated primarily by the medial temporal lobe/hippocampus. On these memory tests, numerous studies have shown that effects of DA on retrieval are minimal.

**Asymmetrical Effects of Divided Attention on Encoding Versus Retrieval**

The existence of an asymmetry of DA effects on encoding and retrieval is unexpected on the basis of traditional theories of memory, which posit that processes implicated at encoding are recapitulated
at retrieval. At the neural level, converging evidence from functional neuroimaging in healthy people and from lesion studies in patients with brain injury also support the idea that the same neural pathways activated when information is perceived and encoded are again reactivated when that information is recovered. Indeed, recent work by Willem Huijbers and his team shows that encoding and retrieval compete for common neural resources. Thus, it is unexpected that DA should create large negative effects on encoding yet produce relatively little effect on retrieval.

Myra Fernandes and Morris Moscovitch helped to clarify this unexpected asymmetry between encoding and retrieval regarding the effects of DA, and in so doing, used the DA technique to specify the component processes critical for retrieval. Their early work showed that substantial effects of DA on retrieval are found when there is overlap between the material used in the concurrent tasks. That is, DA effects at retrieval were said to be material specific. For example, they found that during recall of a list of unrelated words, a visually presented word-monitoring distracting task produced a decrement in memory of 30% from full attention conditions, whereas an equally demanding monitoring task for digit material led to a decrement of only 13%. Unlike the general effect of DA observed at encoding and during strategic retrieval, this material-specific interference effect of DA was interpreted as resulting from competition between the word-based distracting task and the verbal memory test for a common representational system during recovery of the memory trace, and to a negligible extent on competition for general resources. In this sense, the effects of DA at retrieval on long-term memory were similar to those observed at perception and in working memory.

Interference Effects From DA at Retrieval

Subsequent work went on to show that DA effects on word retrieval are similarly observed on memory tests for spatial information. Specifically, Fernandes and Emma Guild examined long-term recognition memory for words or visuospatial patterns under full or DA conditions with a distracting task requiring either phonological (rhyme) or visuospatial (curved line) processing of letters (see Figure 1). They found an interaction such that the curved-line distracting task had a more detrimental effect on memory for spatial patterns than did the rhyme-distracting task, whereas the rhyme-distracting task had a more detrimental effect on memory for words than did the curved-line distracting task. These results show that similarity in processing requirements between the memory and distracting task influences the magnitude of memory interference from DA at retrieval.

Notably, competition for response selection mechanisms under extreme time pressure may also produce large interference effects on retrieval, even for unrelated lists of words regardless of whether the memory task is one that requires strategic processing or not. Mark Carrier, Doug Rohrer, and Hal Pashler showed that when allocation of time to processing the concurrent tasks leaves little time for implementing retrieval processes, retrieval interference is observed. They suggest that memory retrieval cannot be performed under dual-task, or DA, conditions due to response-selection demands of a secondary (distracting) task. Pashler and colleagues showed an increase in response time (RT) on cued recall and recognition when an auditory-manual task was concurrently performed that also required response selection.
Divided Attention and Memory

The increase in memory RT indicated that retrieval was postponed by the response selection task. Thus, according to Pashler’s work, memory retrieval cannot occur in parallel with other cognitive processes and must be postponed until response-related processing resources become available. While this response selection/retrieval bottleneck may contribute to the magnitude of the interference on memory-related responses, it may not implicate memory retrieval processes that precede and guide these responses. The other work described in this entry suggests that the type of material and the type of processing required of the competing, distracting task modulates these retrieval effects.

More recent work is now exploring other scenarios in which memory retrieval is disrupted by DA. Researchers Jason Hicks and Richard Marsh argue that DA will disrupt memory only if control processes are used during retrieval. They showed that recognition of words, presented visually during study/encoding, was unaffected by DA at retrieval, but the hit rate for words initially studied as anagrams (to be effortfully unscrambled by the participant to reveal a word) decreased under DA conditions. The authors argued that the words read intact were more shallowly processed, and, consequently, retrieved by relatively automatic memory processes unaffected by DA during retrieval. In contrast, words generated from anagrams were believed to be more deeply processed at encoding and subsequently retrieved using more recollective-based recognition processes in which memory for the words is accompanied by rich contextual details about the initial encoding episode. Thus, it is this enriched quality of memory that is more susceptible to DA effects at retrieval. Related to this, new research by Erin Skinner and Fernandes shows that recollection draws on strategic processes: Dividing attention during recognition led to errors in recollecting details pertinent to items stored in long-term memory, regardless of the material used in the distracting task. Interestingly, that work also offered support for the material-specificity hypothesis, as familiarity with studied items could be reduced by DA at retrieval when the concurrent and memory task used similar materials.

Conclusions

Dividing attention has proven to be a useful technique not only for showing how attention influences processes in perception, working memory, and long-term memory but also for elucidating the components of the processes themselves. Perhaps the most interesting finding has been the discovery that for some tasks, such as those involving perception and working memory, interference from DA is material-specific, whereas for other tasks, such as encoding of information into long-term memory, interference is material general. Unlike effects from DA at encoding, interference from DA at retrieval varies depending on the demands of the particular memory task, with material-general interference effects for tasks that rely heavily on strategic processes and recollection, but material specific interference for tasks that are based on familiarity and rely less on retrieval strategies.

Myra Annette Fernandes and Morris Moscovitch

See also Attention, Neuroimaging Studies of; Attention, Resource Models; Memory, Interference With; Memory, Neural Basis; Memory Recall, Dynamics

Further Readings


Huijbers, W., Pennartz, C. M., Cabeza, R., & Daselaar, S. (2009). When learning and remembering compete: A
Bayesianism is a movement in epistemology and the philosophy of science that suggests that beliefs come in degrees. Dutch book arguments are a type of argument showing that rational agents must have degrees of belief that obey various principles. These arguments were introduced by Bruno de Finetti to show that for a rational agent, her degrees of belief in various propositions must obey the three axioms of probability theory. That is, every proposition has a nonnegative probability, any proposition the agent is certain of has Probability 1, and the probability of A or B is the sum of the probabilities of A and B, if A and B are incompatible. Dutch books have also been used to establish further constraints on degrees of belief and on how agents should update them over time. The basic form of the argument is to find, for any agent who violates the proposed constraint, a set of bets that the agent finds individually favorable but that collectively would guarantee that the agent loses money—this unfavorable collection of bets is the so-called Dutch book. This entry outlines how Dutch book arguments work, shows how they can be used to support some of the basic tenets of Bayesianism and discusses some of the problems they face.

**Assumptions**

The background of any Dutch book argument is the claim that (beyond the standard propositional attitudes of belief, knowledge, desire, and so on) there is a propositional attitude of degree of belief, and an assumption connecting this degree of belief to the agent’s evaluation of bets as favorable or unfavorable. The assumption is that for every agent and every proposition, there is some number p that counts as her fair price for a bet on that proposition. What this means is that she is willing to pay any amount less than $p for a bet that wins her $1 if A is true, while she will accept any amount greater than $p for a bet that loses her $1 if A is true (or similarly $[100p]$ for a bet that pays $100). These two actions are respectively called buying and selling a bet on A. We typically think of gamblers buying bets and a casino selling them, but the Dutch book argument assumes each agent is willing to do either. This fair price p is then said to be the agent’s degree of belief in A.

For example, let’s say your degree of belief is .6 that it will rain tomorrow in Seattle. The argument assumes that you will be willing to pay $59.99 to place a bet where you win $100 if it rains tomorrow in Seattle, but you will not be willing to pay $60.01. However, you will be willing to accept if someone else pays you $60.01 to place a bet where you have to give them $100 if it rains tomorrow in Seattle. This is clearly an idealization—in some cases there will be a range of values at which you will be unwilling to either buy or sell, while in other cases there may be a range at which you are happy to do either. This is connected to the phenomenon of risk aversion.

**The Basic Argument**

But if we do accept this idealization, then the Dutch book argument for the three probability axioms is straightforward. An agent’s fair price for any bet should never be negative, because then she’d be willing to give away some money to sell a bet that could only make her give away more money. If the agent is certain that A is true, then her price should be 1. If it’s lower, then she’s certain she would lose more than she’d be willing to sell the bet for, and if higher, then she’d be willing to buy it for more than she could win. Finally, if A and B are incompatible, then her price for A or B should be the sum of her prices for A and B. To see this, consider an example. If your degree of belief that Speedy will win the horse race is .2, your degree of belief that Dasher will win is .3, and your degree of belief that either Speedy or Dasher will win is .4, then someone can be sure to make money off you by making three bets. First, you pay them $19.99 for a bet that wins you $100 if Speedy wins. Then you pay $29.99 for a bet that
wins you $100 if Dasher wins. Then they pay you $40.01 for a bet that costs you $100 if either Speedy or Dasher wins. At this point you have lost $9.97. But once the race is over, if either Speedy or Dasher wins, then you each pay each other $100, while no more money changes hands if a different horse wins. Thus, no matter what happens, you lose $9.97. If your price for “either Speedy or Dasher will win” had been .5 (the sum of your individual prices), then the person could not have made money from you.

Worries About the Argument

One worry about this argument is that it only shows that an agent whose betting behavior violates the probability axioms has financial problems and not that she has epistemic problems. But Brian Skyrms suggests that the Dutch book actually shows that the agent values exactly the same situation in two different ways and is thereby incoherent. For the first two cases, this is quite clear, because she must view the situation both as a bet that she finds favorable and also as a situation in which she is guaranteed to lose money. The third argument raises further issues because it involves multiple bets.

The Skyrmsian interpretation of the Dutch book argument also gives a way to avoid the problem of risk aversion. Although there may not be any specific price that divides buying and selling as required for the argument, David Christensen suggests that an agent’s degrees of belief must “sanction a price as fair,” even though the agent’s risk aversion prevents him from buying or selling in a range around that price. Although an agent whose degrees of belief violate the probability axioms may (if he’s risk averse) never run into financial problems, he still implicitly evaluates a set of bets as fair even though they guarantee a loss.

One further difficulty with the interpretation of Dutch book arguments is understanding what it means to “guarantee” a loss. The argument above interprets this as meaning that the agent herself is certain that there would be a loss. Another popular interpretation is that the guarantee must be a logical guarantee of a loss—note that an agent can be certain of be 1, the logical interpretation requires that an agent’s degree of belief in any logical truth be 1. Different interpretations of guarantee will give rise to conclusions with different versions of the axioms for probability. The standard axioms (due to Andrey Kolmogorov) describe a probability function as defined on a collection of abstract sets. The interpretation suggested here interprets the elements of these sets as epistemic possibilities for the agent on which the proposition is true, while the logical interpretation takes the elements to be logical possibilities on which the proposition is true.

Conclusion

Although there are several worries about Dutch book arguments, the basic conclusion that degrees of belief must obey the probability axioms (in some form or other) has been supported by various other arguments as well. However, there has been less independent support for the conclusions of other Dutch book arguments. (For those who are familiar with the notions, Dutch books have been used to support versions of the additivity axiom with infinitely many disjuncts rather than just two, as well as the Bayesian rule of learning by conditionalization, and many other supposed requirements of rationality for belief.)

The worries about Dutch books have helped clarify the nature of degrees of belief and the constraints they ought to satisfy for rational agents. In the basic cases, these arguments are not the only arguments for the same conclusion, so the worries aren’t too troublesome. But in some cases involving infinity or the change of beliefs in response to evidence, much work remains to be done to show that these arguments really support the claimed conclusions.

Kenny Easwaran

See also Allais Paradox; Belief and Judgment; Decision Theory, Philosophical Perspectives

Further Readings


Dyslexia means “impaired reading.” Dyslexia can arise in two ways. A person may fail to learn to read competently (developmental dyslexia) or a person may have become completely competent at reading and then lose some of this reading ability because of brain damage (acquired dyslexia). Acquired dyslexia (like developmental dyslexia) occurs in a variety of forms that differ from each other with respect to the particular types of reading symptoms the patient shows. This entry discusses the six clearly distinguishable forms of acquired dyslexia and also mentions acquired dysgraphia (spelling impairment following brain damage).

### Pure Alexia

Pure alexia, also known as alexia without agraphia (dysgraphia), or letter-by-letter reading, is the only form of acquired dyslexia in which spelling can be preserved (hence, the “pure” and the “without agraphia”). When the patient is unable to read a word promptly, it is typically spelled aloud by letter from left to right (hence, the third name for this disorder); if all the letters are named correctly, the patient can usually then produce the whole word. It is not clear how this acquired dyslexia arises, but a promising current theory is that it is due to brain damage specifically harming the process of letter identification. Letters such as L or Z, which are not very confusable with other letters, can still be read easily, but letters such as E or Q, which are highly confusable with other letters, are very difficult to distinguish from the letters they are confusable with. To guarantee correct letter identification here, the patient must focus all attention on the letter rather than, as normal readers do, distribute attention across all the letters in the word. So attention has to be devoted to single letters, letter by letter, and that’s what causes the letter-by-letter reading behavior.

### Surface Dyslexia

Irregular words are those that disobey standard letter-sound rules, words such as *yacht* or *blood*. Patients with surface dyslexia frequently misread such words, and, what is more, they make “regularization errors”: reading the irregular words according to the rules (so *yacht* is read as if it rhymed with *matched* and *blood* as if it rhymed with *mood*). Accuracy of reading of regular words and of nonwords can be normal. This is typically interpreted in terms of a dual-route model of reading aloud. In such a model of reading, one route from print to speech uses letter-sound rules: This route (the “nonlexical route”) correctly reads regular words and nonwords but makes regularization errors with irregular words. The other reading route, the “lexical route,” operates by recognizing real words as familiar wholes: It correctly reads regular and irregular words but cannot read nonwords (such as VIB or SINT) since nonwords are not familiar wholes. Given such a model of reading aloud, surface dyslexia is ascribed to an impairment of the lexical route for reading aloud; the nonlexical route may be entirely intact.

### Phonological Dyslexia

Here, there is a selective impairment in the reading aloud of nonwords. Accuracy of reading aloud regular and irregular words is much better and may even be intact. This form of acquired dyslexia is also usually interpreted in terms of the dual-route model of reading aloud: The nonlexical reading route is selectively damaged, with the lexical reading route relatively spared or even intact.

### Deep Dyslexia

The key symptom of this acquired dyslexia is the semantic error in reading aloud. The patient might read *canary* as *parrot* or *antique* as *vase*. This is the most striking symptom, but all deep dyslexics also show several other symptoms: They are worse at reading aloud function words such as *from* or *but* than content words such as *form* or *bet*, and worse at reading aloud abstract words such as *form* or *bet* than concrete words such as *farm* or *bat*. Nonword reading is impossible. Deep dyslexia appears to differ from the other five acquired dyslexias in the following way: For the other five, the
patient is reading using a damaged version of the normal reading system (which is located in the left hemisphere of the brain), whereas the deep dyslexic is unable to make any use of that system and is instead reading using a rudimentary reading system located in the right hemisphere of the brain, a system that may play little or no role in reading by people with intact brains.

Word-Meaning Blindness

Here, printed words can be recognized (the patient can tell that isle is a real word, whereas sint isn’t) and can be read aloud, even highly irregular words, but reading comprehension is impaired. This is not because the patient has lost knowledge of word meanings, because comprehension of spoken words is intact. So this acquired dyslexia is interpreted as due to a disconnection of the (intact) visual word recognition system from the (intact) systems of word meanings.

Lexical Nonsemantic Reading

As in word-meaning blindness, printed words can be recognized (the patient can tell that isle is a real word, whereas sint isn’t) and can be read aloud, even highly irregular words, but reading comprehension is impaired. However, in contrast to word-meaning blindness, this failure of reading comprehension is due to loss of knowledge of word meanings, because the patient is just as bad at understanding spoken words or pictures as at understanding printed words.

Brain damage can also impair spelling, and acquired dysgraphia, like acquired dyslexia, also comes in different forms. These include surface dysgraphia (selective difficulty in spelling words that disobey standard spelling-to-sound rules), phonological dysgraphia (selective difficulty in spelling nonwords), and deep dysgraphia (semantic errors in writing to dictation, such as writing canary when asked to write parrot). Acquired dysgraphias are usually interpreted in terms of a dual-route model of spelling.

Max Coltheart

Further Readings


Dyslexia, Developmental

This entry considers the cognitive and sensory characteristics of dyslexia as well as current neural and genetic associations with this disorder of learning. Developmental dyslexia is usually characterized as a specific problem with reading and spelling that cannot be accounted for by low intelligence, poor educational opportunities, or obvious sensory or neurological damage. Because children with developmental dyslexia frequently have good spoken language skills, the earliest theories of developmental dyslexia were visual rather than linguistic. Developmental dyslexia was conceived as a form of “word blindness.” Today, developmental dyslexia is recognized as a primarily linguistic disorder. Children with developmental dyslexia have difficulty in the neural specification of the sound structure of language—a specific difficulty in phonological representation that has been found in every language studied to date.

The Development of Phonological Representation

Child phonology has undergone dramatic theoretical revision in recent years. The traditional view of phonological development was that babies began to learn language by recognizing phonemes. Phonemes are the individual sound elements that appear to
make up words in languages and that roughly correspond to alphabetic letters. It was believed that all languages drew on a universal phonetic inventory of consonant and vowel phonemes. Linguists such as Janet Pierrehumbert proposed that babies actually learn language-specific “phonotactic templates” or “prosodic structures.” These templates are essentially phonological patterns that vary in sound intensity, pitch, duration, and rhythm. A common template for English is a bisyllabic pattern, with stronger first-syllable stress (a strong—weak stress template). The strong first syllable is typically louder, longer and higher in pitch than the second syllable. Familiar words that follow this pattern are *mummy, daddy, biscuit,* and *baby.*

As demonstrated in the cross-language review of reading development and dyslexia by Jo Ziegler and Usha Goswami, phonological representation in a preliterate child is different from phonological representation in a child who has achieved literacy. This is because learning to read changes the brain. Once the alphabetic code has been acquired, spoken language processing changes fundamentally. The literate brain imposes phonemes onto the sound structure of speech. The preliterate brain does not. Adults who have never learned to read perform very poorly in phoneme awareness tasks such as phoneme deletion (e.g., deleting the phoneme /t/ in *stop* to leave *sop*). Both preliterate children and illiterate adults perform well in phonological awareness tasks based on larger phonological units, however, such as syllables and rhymes. Hence, phonological representation prior to the teaching of literacy appears to be based on syllables and the subsyllabic segments of onset (any sounds before the vowel) and rime (the vowel and any sounds that follow it). The onset-rime segmentation of *sing, sting,* and *spring* would be *s-ing, st-ing* and *spr-ing,* respectively.

### The Dyslexic Phenotype

In developmental dyslexia, there is a brain-based difficulty in representing the sound structure of speech. This selective difficulty with phonology is usually characterized by extremely poor performance in three areas of phonological processing. The first is phonological awareness of syllables, rhymes, and phonemes. For example, children with developmental dyslexia have difficulty in counting the syllables in a word such as *university*; they have difficulty in selecting the odd word out from a triple such as *man, pat, fan,* and they have difficulty in substituting phonemes in Spoonerism tasks (e.g., *Bob Dylan* becomes *Dob Bylan*). The second is phonological memory skills. Children with dyslexia find it difficult to remember verbal sequences and show poor performance on standardized tests of phonological memory such as nonword repetition (recalling items such as *loddernapish* and *thickery*). The third is *rapid automatized naming* (RAN). Children with dyslexia find it difficult to rapidly produce highly familiar phonological forms such as color names, object names, digits, and letters. These three areas of phonological processing are usually all impaired in a child who has been identified as having developmental dyslexia.

### Sensory-Processing Problems?

A number of sensory-processing theories have attempted to explain these hallmark difficulties with phonology. Representative theories include theories postulating auditory-processing deficits linked to temporal cues such as Paula Tallal’s rapid auditory processing deficit theory and Usha Goswami and colleagues’ slower temporal modulations theory. John Stein’s theory (with Vincent Walsh) postulates visual-processing deficits in the magnocellular system, while Mark Seidenberg and his colleagues discuss difficulties with noise exclusion in perceptual tasks, and Andrea Facoetti and colleagues suggest difficulties with multisensory spatial attention. Currently, theories based on auditory temporal processing currently show the greatest convergence across languages. However, it is logically possible that there may be a variety of causes of dyslexia, with some children’s primary sensory problems lying outside the auditory realm. It is also logically possible that a primary difficulty in the auditory system may lead developmentally to atypical development in other systems, such as visual or attentional systems. This would explain the wide nature of sensory deficits that are typically found when relatively old dyslexic children are tested.

### Brain-Imaging Studies

Current brain-imaging data support the consensus view that the primary weakness in dyslexia involves the representation of phonological information. This primary weakness does not disappear with age, and
although phonological treatments can be effective in promoting reading achievement, even well-compensated dyslexic adults show phonological problems and read very slowly. Most studies reveal atypical brain activation in the three important sites for reading: the left posterior temporal regions, the left inferior frontal regions, and the left occipitotemporal regions. These sites suggest problems with both the phonological aspects of reading and with the efficient development of a mental dictionary of word spellings—an “orthographic lexicon.” The typical underactivation in the core left temporoparietal networks can be normalized with phonological training, and there is evidence that articulatory networks may play a compensatory role.

Genetic Studies

Developmental dyslexia is highly heritable, with linkage studies by Elena Grigorenko and colleagues demonstrating important sites on the short arm of Chromosome 6 and on Chromosome 15. Such studies depend on a dyslexic phenotype defined on the basis of a phonological deficit. Genes for cognitive disorders are not deterministic, because reading is a culturally determined activity. Therefore, individuals at genetic risk for dyslexia who develop in a favorable early environment (e.g., children with caregivers who stimulate their linguistic awareness) may be able to compensate for their genetic predisposition. However, levels of association reported so far in behavioral and molecular genetics are not strong enough to translate into reliable predictors of risk for a single individual. Neural markers may offer an alternative for the early identification and remediation of dyslexia.

Usba Goswami

See also Dyslexia, Acquired; Dyslexia, Phonological Processing in; Speech Perception; Word Recognition, Auditory; Word Recognition, Visual

Further Readings


Dyslexia, Phonological Processing in

Developmental dyslexia is by definition a disorder of written language acquisition, despite adequate intelligence and opportunity, and in the absence of obvious sensory, neurological, or psychiatric disorder. It primarily affects the acquisition of reading, particularly word identification and, secondarily, the acquisition of conventional spelling.

Underlying Causes of Developmental Dyslexia

Two main proximal causes have been considered. Historically, the initial hypothesis was that of a visual deficit (“congenital word blindness,” coined by William Pringle-Morgan in 1896). In the 1970s, it became evident that what had been interpreted as visual letter confusions were better explained by phonological confusions. Over the last three decades, it has been well established that most cases of dyslexia can be attributed to a subtle disorder
of oral language (the “phonological deficit”), with symptoms that happen to surface most prominently in reading acquisition. It remains likely that a minority of cases of dyslexia are due to disorders in the visual modality, although the precise nature of the deficit remains unclear. The present entry focuses on cases of dyslexia with a phonological deficit.

Another important theoretical debate is whether the phonological deficit in dyslexia is specific to the linguistic domain or is caused by an underlying auditory deficit. Although there is considerable evidence that a subset of dyslexic children have difficulties in a variety of auditory tasks, there have been important challenges to the view that this is the underlying cause of their phonological deficit, hence the cause of their reading disability. Again, given that both sides of the debate agree that the phonological deficit is central to understanding dyslexia, this issue will not be further discussed here.

**Symptoms of the Phonological Deficit**

There is wide agreement on the main symptoms of the phonological deficit in dyslexia: They include poor phonological awareness, poor verbal short-term memory, and slow lexical retrieval. Phonological awareness refers to the realization that words are made of a combination of smaller units (syllables and phonemes) and to the ability to pay attention to these units and explicitly manipulate them. Typical tasks include counting the number of syllables or phonemes in a word, detecting whether words rhyme, deleting the initial (or final) phoneme, or performing simple spoonerisms (swapping the initial phonemes of two words). Verbal short-term memory typically refers to the ability to retain and immediately repeat verbal material of increasing length: sequences of two to nine digits (digit span), nonwords of two to five syllables (nonword repetition), or even sequences of nonwords (nonword span). Finally, lexical retrieval refers to the ability to quickly retrieve the phonological forms of words from long-term memory. In the context of dyslexia research, this is tested by having participants name series of 50 objects, colors, or digits as fast as possible (rapid automatized naming). Dyslexic children are typically found to have poor phonological awareness (particularly phoneme awareness), reduced short-term memory span, and slow automatized naming, although individual profiles along those three dimensions of course vary, leading to the possibility of subtypes. Thus, the most prominent symptoms of developmental dyslexia are diverse but united by their involvement of phonological representations—hence, the consensus hypothesis of a phonological deficit.

**Nature of the Phonological Deficit**

The study of the symptoms of dyslexia has led many researchers to hypothesize that dyslexics’ phonological representations are somewhat degraded, poorly specified, noisy, lacking either in temporal or in spectral resolution, or are insufficiently attuned to the categories of the native language. An alternative view is that phonological representations in dyslexia are intrinsically normal but that the observed difficulties in certain (but not all) phonological tasks arise from a deficit in the access to these representations, a process that is particularly recruited for short-term memory, speeded retrieval, and conscious manipulations. The precise nature of the phonological deficit therefore remains to be uncovered.

**Consequences of the Phonological Deficit**

Beyond the observation that most dyslexic children have some form of phonological deficit, the hypothesis is of course that this deficit is the direct cause of the reading disability. In particular, phonological awareness is seen as a major cognitive prerequisite for the acquisition of the mappings between graphemes (letters or groups of letters) and phonemes, which themselves provide the foundation of reading acquisition. Verbal short-term memory and efficient lexical retrieval are also thought to play a role in reading acquisition. Indeed, there is ample evidence that these phonological skills are not only defective in dyslexic children but more generally predict reading ability, both in dyslexic and in normally developing children. However, it has been shown that reading acquisition itself improves phonological skills so that correlations between the two dimensions do not unambiguously indicate the direction of causation. The differences in phonological skills between dyslexic and control children could similarly be interpreted as resulting from their different levels of reading ability. Indeed, illiterate adults who didn’t have an opportunity to learn to read have also been found to show poor phonological skills.
More definitive evidence for the causal connection between poor phonological skills (and in particular poor phonological awareness) and dyslexia has come from several additional lines of enquiry. In many studies, dyslexic children have been shown to have poorer phonological skills, not only than normal readers of the same age but also than younger children who have the same reading level. Furthermore, longitudinal studies beginning before reading acquisition have established that phonological skills predict reading ability several years ahead and that the phonological deficit is present in would-be dyslexic children even before they learn to read. Finally, a few longitudinal studies starting from birth, using both behavioral methods and event-related potentials, have established that the phonological deficit may have precursors already in the first year of life, in the form of poor categorization or discrimination of speech sounds.

In summary, there is overwhelming evidence that a phonological deficit is a proximal cause of reading disability in at least a majority of dyslexic children. The precise nature of this deficit remains to be fully understood, and its neural and genetic bases are also under intense scrutiny.

Franck Ramus

See also Dyslexia, Developmental; Eye Movements During Reading; Language Development; Word Learning; Word Recognition, Auditory

Further Readings


Intracellular neural currents carry inhibition and excitation from the synapses where neurons communicate to where they decide whether to fire. Those same currents give rise directly to the magnetic fields measured by the magnetoencephalogram (MEG); the equal but opposite extracellular return currents instantaneously cause potential differences between locations on the scalp, which are measured by the electroencephalogram (EEG) (Figure 1). The EEG and MEG thus directly and instantaneously reflect the computational currents used by the information-processing mechanisms of the human brain.

Consequently, the EEG and MEG have a systematic relationship to different aspects of mental activity, including higher mental functions such as attention, orienting, memory, language, face perception, action, error correction, reanalysis, and closure. Typically, these relationships are revealed by averaging the EEG or MEG with respect to repeated cognitive events, yielding event-related potentials (ERP) and event-related fields (ERF), or generically, evoked responses (ER). Across thousands of studies, it has been observed that ERs can be described as a series of positive and negative peaks with characteristic latencies, scalp topographies, and cognitive correlates. These regularly occurring clusters of mental and physical characteristics have been reified as named components. This approach provides a vocabulary for communicating discoveries in EEG and MEG and thus bootstraps a physiological lexicon of the mind.

Because they measure directly and instantaneously the currents that perform information processing in the human brain, EEG and MEG have a temporal and physiological accuracy that hemodynamic measures, such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI), lack. However, EEG and MEG have other problems. First, it is impossible to localize with certainty from extracranial measures where they are generated in the brain. This not only renders problematic the interpretation of ER with respect to neural systems; it also deprives components of an unambiguous defining characteristic and thus makes component-based analysis somewhat circular. Second, most of the information-processing currents cannot be recorded at a distance because of spatiotemporal cancellation. Action potentials are not recorded at all, nor are currents from interneurons. EEG and MEG arise mainly from relatively slow and synchronized currents in the apical dendrites of pyramidal cells. Third, the polarity of the EEG or MEG signal does not tell us whether the underlying neural activity is excitatory or inhibitory. To some degree, these difficulties can be ameliorated by combining information from PET or fMRI with that of EEG and MEG, within the context of individual brain anatomy provided by structural MRI. In certain clinical situations, it is possible to record potentials directly within the human brain, yielding very high spatial, temporal, and physiological accuracy; such measures can be used to infer the location
and physiological meaning of EEG and MEG components recorded in healthy subjects.

Unlike hemodynamic measures that allow a modern physiological phrenology correlating brain locations with mental functions, cognitive ER components are best thought of as embodying distributed cortical information processing modes. It seems that there are only a limited number of such modes and that they occur not only during ER but also during the spontaneous brain activity associated with sleep and waking, when they can be recognized as named EEG or MEG rhythms. In many cases, ER result from stimulus-evoked, phase locking of these rhythms.

Rhythms and Their Brain States

Alpha

The earliest observations of EEG noted the presence of a prominent posterior rhythm at about 10 hertz (Hz) that is suppressed by mental effort or opening the eyes. This alpha rhythm in the visual modality has a counterpart in the somatosensory and motor domain termed the mu rhythm, which is suppressed by movement and extends to frequencies around 20 Hz. A homologous rhythm in the auditory cortex named the tau rhythm is suppressed by auditory stimulation and is more visible in MEG than EEG. More generally, rhythms in this domain are suppressed by both heightened alertness and by drowsiness. They seem to indicate a state of readiness in the generating cortex and may modulate information processing.

It is hypothesized that all parts of the cortex can generate alpha. The neural circuitry underlying the alpha is not known but is thought to be the same as that of the sleep spindle. The sleep spindle consists of a burst of 10 to 16 Hz activity lasting about a second, occurring hundreds of times every night during Stage 2 of non-rapid eye movement (NREM) sleep (i.e., sleep without vivid, elaborated dreams). The function of sleep spindles is not known, but they have been implicated in declarative memory consolidation. When thalamic cells are hyperpolarized during sleep, intrinsic membrane currents are activated that support burst-suppression firing at alpha frequency. This rhythmic firing is reinforced by an internal thalamic network between the thalamocortical projection cells and neurons of the thalamic reticular nucleus. Alpha frequencies are projected from the thalamus to the cortex, but the return corticothalamic projection is necessary to synchronize the alpha across the thalamocortical network.
There are two thalamocortical systems. The core system relays specific information; the matrix system synchronizes cortical domains. Each core thalamocortical neuron makes strong synapses in narrowly focused cortical areas; each matrix neuron makes modulatory synapses in widespread cortical areas. It appears that MEG spindles reflect mainly the core system, whereas EEG spindles reflect mainly the matrix system. Current evidence suggests that the core thalamocortical system generates the alpha, but the matrix system may also contribute.

**Delta**

During deeper stages (3 and 4) of NREM sleep, the brain alternates every second between near electrical silence because of inward potassium currents, called down-states, and cortico-cortical activity levels close to those of active waking, or up-states. This alternation is known in animals as the slow oscillation and in humans as delta or slow wave activity. The function of the slow oscillation is again unknown, but the repeated rebooting of consistent sequences of cortical neuronal firing during the transition from the down-state to the up-state has been hypothesized to consolidate their continued association in waking. During Stage 2 NREM sleep, an isolated down-state termed the K-complex can be evoked by sensory stimuli that do not wake up the subject.

**Theta**

In felines and rodents, the theta rhythm, around 8 Hz, is most prominent in the hippocampus during instrumental movement and alert attention. Theta modulates neuronal firing and gamma activity, and theta phase is related to movement of the animal through its allocentric cognitive map. In humans, hippocampal theta is less prominent and more difficult to distinguish from pathological activity. However, intracranial recordings do support its presence and its relationship with theta in association cortex, especially during declarative memory retrieval. It may also be related to sustained working memory operations, where a frontal midline theta is evoked. The cortical theta comprises synaptic currents rhythmically alternating between the middle and upper cortical layers. Middle cortical layers receive inputs from lower level sensory cortex and specific thalamic relays. Upper cortical layers receive inputs from higher associative cortex and modulatory thalamic systems. Thus, the human cortical theta may represent an alternating predominance of external (sensory) and internal (associative) influences on local information processing.

**Gamma**

It is important to distinguish between two kinds of gamma activity. Gamma oscillations are confined to a fairly narrow frequency band that, for any given instance, may be somewhere between 30 Hz and 80 Hz (or even up to 200 Hz in pathological situations). Gamma oscillations seem to arise from the synchronous firing of networks of inhibitory interneurons. Primary sensory cortices are especially prone to generating such oscillations, which in animals may provide envelopes of cortico-cortical interaction and have been hypothesized to play a role in binding (the creation of transcortical, resonant cell assemblies embodying the various aspects of mental events).

The other kind of gamma activity is broadband. It extends from less than 10 to over 1,000 Hz, and its power declines approximately as the square of the frequency. It represents integrated synaptic activity, with increasing contributions from action potentials at higher frequencies. Thus, it is a faithful measure of the amount of local neural activation and is strongly correlated with fMRI activation.

Special caution should be taken in gamma recordings to eliminate contamination from non-neural sources. Gamma power can be over 1,000 times weaker than power in low frequencies. Muscle activity or micromovements of the eyes are in the same frequency range as gamma but often have much larger amplitude.

**Components**

**A Matter of Definition**

ERP components are commonly defined by their topography, latency, and cognitive correlates. For example, the P3 (or P300) is commonly defined as a positive potential recorded at the top of the head, peaking at about 300 ms (milliseconds) after stimulus onset, and larger to rare events. However, this definition is inadequate because the P3 can be recorded with other measurement modalities, at other locations, at other latencies, and in other cognitive circumstances. Furthermore, it is clear that the
scalp P3 actually is a superposition of different brain states, and conversely, that brain states not associated with the P3 can contribute to scalp potentials at 300 ms latency. To avoid these difficulties, this entry will consider components as particular brain states rather than as shorthand for particular sensorvalues. When referring to the measurement, a prefix will be added to the component name indicating the recording location or modality.

**Orienting and Closure: P3 and Related Components**

The P3 indexes at least three generic neurocognitive brain states. One state continues earlier modality-specific processing in the auditory, visual, and somatosensory association cortices. A second, the P3a, is related to the orientation of attention. A third state, embodied in the P3b, updates and closes cognitive contextual integration (see below). Typically, the scalp P3 reflects the superposition of the P3a, which peaks earlier, and the P3b, which peaks later.

The P3a is evoked by rare target and distracter stimuli, regardless of whether they are overtly attended. It is only weakly evoked by abstract semantic operations, but instead is evoked by stimuli that demand processing because of their potential biological significance. It is not modality specific but is more easily evoked by auditory than visual stimuli, presumably reflecting their greater ability to elicit orienting. Such stimuli also evoke the *orienting response*, typically measured as various autonomic phenomena, including the skin conductance response, which can be highly correlated with the P3a.

The P3a is generated in areas that support spatial attention, including the dorsolateral, prefrontal, and supramarginal cortices, as well as the cingulate and other paralimbic cortices. Additional weaker diffuse activation may represent the widespread polling of cortical areas to arrive at a rapid evaluation of the stimulus. The frequency content of the triphasic sharp waveform that contains the P3a, its cortical distribution, and its close relationship to attention suggest that it may be generated by the thalamocortical circuits that generate alpha and spindles (see above).

The P3b is evoked by stimuli in all modalities and even by the absence of an expected stimulus. It only occurs to attended, task-relevant stimuli. Indeed, the P3b is present if, and only if, independent behavioral data show that the stimulus has reached awareness and been subjected to controlled processing. P3b onset occurs at about the same latency as the specification of the subject’s response, suggesting that the P3b begins when event-encoding activities are substantially complete and the brain is ready to move on to its next task, updating the contents and instructions in the mental workspace as well as entering the conclusions of the event-encoding process into working and declarative memories.

These activities are centered in the interaction of ventral temporofrontal event-encoding cortices (inferotemporal, perirhinal, and ventrolateral prefrontal), association cortices (superior temporal sulcal and posterior parietal), and the hippocampus, where the P3b is generated. Updating may reflect top-down associative circuits that activate GABAb and NMDA receptors in upper cortical layers shared with the theta described above. Closure may be implemented with voltage-gated potassium channels, as in cortical down-states. Cholinergic and monoaminergic projections are likely to act indirectly on thalamus and cortex to encourage the different P3 states.

The P3s are preceded by various components that generally do not require attention and are modality specific. Notable among these is the mismatch negativity (MMN) evoked by deviant stimuli in a rapid auditory stream and used extensively to study the formation and contents of preconscious auditory memory. Spatial and cross-modal selective attention both modulates early components and superimposes a negativity termed the *N2pc*. Although controversial, most evidence indicates that the earliest components in primary cortex are not modulated. When the P3b terminates a sustained cognitive context, that context is often marked by a negativity, the contingent negative variation (CNV), which may reflect sustained cell firing supporting the maintenance of items in the mental scratchpad.

**Material Specific Encoding, Semantic Associations, Syntax, and Reprocessing**

Semantically meaningful stimuli, symbols that provide an entry point to rich learned associations, evoke a characteristic series of components embodying first, template matching, then semantic associations, and finally, if necessary, reanalysis. Template-matching components have been mainly studied in the visual modality where they peak at approximately 170 ms in small specialized patches
of the posterior fusiform gyrus: the visual word form area on the left and the fusiform face area mainly on the right. The N170 (N indicates that it’s a negative-going voltage peak) encompasses bursts of units firing gamma, which are projected to the middle layers of widespread temporofrontal association areas at approximately 200 ms.

This material-specific template matching determines if the stimulus is potentially symbolic or semantic, and if so, encodes it and triggers an N400. The N400 appears to reflect associative cortico-cortical synaptic currents in upper cortical layers of classical language areas (Wernicke’s and Broca’s) as well as the ventral temporal language area, with weaker activity in contralateral and occipitotemporal sites. Neuronal firing specific for individual words, people, or semantic categories occur in these areas during the N400. N400s to visual or auditory words are generated in mainly overlapping cortical areas. Once evoked, N400 amplitude is decreased if contextual information facilitates integration of the stimulus with the cognitive context. Modulation is effective across different sensory modalities, memory systems, and semantic domains, suggesting that the N400 embodies an abstract network interrelating these diverse realms.

Although semantically deep, the N400 provides rapid associations that may lead to erroneous interpretations. When this happens—for example, during garden-path sentences and puns—a P600 is evoked, engaging cingulate and right dorsolateral prefrontal as well as language cortices. Complex syntactic anomalies can also evoke a P600, whereas word category or phase structure violations can evoke an early left anterior negativity (ELAN) at approximately 200 ms.

Response Generation and Monitoring

Prior to movement, perhaps prior even to the intention to move, a slowly rising negativity termed the readiness potential (RP) or bereitschaftspotential (BP) is generated in premotor and supplementary motor cortices. Just before the movement, the RP begins to rise sharply over the active motor cortex, producing the lateralized readiness potential (LRP). If the response has been in error, a sharp error-related negativity (ERN) peaking approximately 100 ms after response onset is generated in anterior cingulate and related cortices. Other components, related to the attachment of morphosyntactic markers to words and their articulation, are generated in Broca’s area prior to speech, but they have not been studied extracranially.

Conclusion

This entry has first reviewed the named EEG or MEG rhythms and their associated brain states, and then has discussed the cognitive correlates and neural generators of the major cognitive ER components. These phenomena provide a lexicon for the physiology of mind—a means to relate synaptic activity, neural circuits, and cortical areas to cognitive information processing.

Eric Halgren

See also Attention, Neuroimaging Studies of; Sleep and Dreams

Further Readings


ELIMINATIVE MATERIALISM

This philosophical doctrine is simultaneously optimistic and pessimistic. First, it is decidedly optimistic that the explanatory resources of psychology and the several neurosciences will eventually provide an exhaustive (physical) account of all mental phenomena. In this modest respect, the view lines up with the more familiar forms of philosophical materialism, such as reductive materialism (i.e., the mind-brain identity theory) and functionalism. Second, it is sharply distinguished from these views by being pessimistic that the scientific account thereby provided will preserve, vindicate, or explain the familiar
categories so central to our current commonsense or folk psychology—categories such as believes that P, desires that Q, perceives that R, decides that S, and so forth, where the variables P, Q, R, and so on, stand in for some declarative sentence or other. Instead, the eliminative materialist expects those categories, so central to our everyday explanatory and predictive practices, to be superseded by, and eventually eliminated in favor of, a new set of categories provided by a successful scientific account of the “real” kinematics and dynamics of brain activity. The central claim is that the propositional attitudes—as philosophers have come to call them—are not the fundamental elements of cognition, nor are they the true causal determinants of human and animal behavior. In fact, the claim continues, strictly speaking, they don’t even exist, despite the assumptions of common sense. In time, then, the vocabulary that purports to describe them will eventually be eliminated from our explanatory practices.

Proposed Illustrative Parallels

The sort of conceptual revolution here contemplated is said to have numerous instances in our intellectual history. For example, in the 17th and 18th centuries, classical thermodynamics spoke systematically of a fluid substance called caloric, supposed to be responsible for all thermal phenomena. Caloric was said to flow from body to body, was constrained under pressure, participated in chemical reactions, and made steam engines go—it was thought. But despite its nontrivial explanatory and predictive virtues, caloric theory became plagued with explanatory and predictive failures, and it was eventually displaced, in the late 1800s, by statistical mechanics, a much superior theory that identified heat with the motions of submicroscopic molecules rather than with a macroscopic fluid substance. Caloric theory simply could not be squared with this new account of the reality underlying thermal phenomena. Caloric fluid was therefore eliminated from our scientific ontology and from our explanatory practices. The micromechanical framework simply took its place.

A second alleged parallel concerns the alchemical substance phlogiston, long supposed to be the principal element released into the atmosphere during the burning of any combustible substance or the rusting of any metal. This pre-Lavoisieran theory displayed, again, nontrivial explanatory virtues, but it was eventually displaced by the oxidation theory of combustion and rusting, a theory that said both processes involved not the release of something, but the ingestion of something: oxygen. Phlogiston was thus eliminated from our ontology entirely and was replaced by the various elements, compounds, and transformations of Lavoisier’s new chemistry.

These cases exemplify the sort of conceptual revolution anticipated by the eliminative materialist, but a standard objection is that they are drawn from the theoretical stratosphere rather than from the realm of commonsense observables. Can we really assimilate the “manifest” states of our own consciousness to states that are plainly theoretical? The eliminativist replies that people at the time took caloric fluid flow to be eminently observable. Place a warm stone (or a snowball) in your hand: The inflow (or outflow) of caloric was manifest, to the touch, to anyone who possessed the concept. Even the evanescent phlogiston could be “seen” leaping skyward from any fire.

These historical examples are all very well, it is commonly objected, but why should we expect that a similar fate awaits the various propositional attitudes, those representational and computational vehicles so central to our folk psychological descriptions and explanations? What comparatively devastating defects allegedly loom that portend a comparably revolutionary fate?

The Presumed Defects in Our Current Folk Psychology

Here the eliminative materialist alleges a number of problems. The first worry concerns our “folk psychology’s” essential reliance on the structural elements of human language—namely, the endless varieties of declarative sentences that give the propositional attitudes both their representational content and their inferential roles within cognition. The problem is said to be that only humans display a command of those linguistic structures. No other creature seems able to learn them. And yet nonhuman animals are no less cognitive creatures than we are. They plainly learn, perceive, solve complex problems, anticipate the future, and display highly intelligent behavior. But whatever is going on inside their heads would seem not to be a dance of sentence-like states, as we commonly suppose to be
taking place inside ours. In sum, it is complained, we humans seem to have seized upon an idiosyncratic form of interpersonal communication, unique to our own species, as the model for conceiving the basic elements and operations of cognitive activity for all the creatures on the planet.

A second worry derives from our growing scientific understanding of how the microstructure of brains generally, ours included, actually does support the activities of representation and computation. Experiments in vivo and computer models of artificial neural networks both indicate that any given brain area represents a fleeting reality with a fleeting pattern of activation levels across its entire neuronal population. This is called population coding or high-dimensional vector coding, and given the number of neurons typically involved (in the millions), it is extraordinarily powerful. It also has nothing to do with sentences.

The same experiments indicate that the brain computes (i.e., performs transformations on its representations) by sending the activation pattern of one population of neurons through a matrix of synapses that connect that population to a second, receiving population. Those many millions of synapses effect a principled transformation of the initial activation pattern into a new and different pattern now displayed across the second neuronal population. This is called parallel distributed processing, and it displays prodigious computational powers, for it performs many millions of elementary computations (one at each participating synapse) simultaneously and all at once. It is arguably a form of inference, perhaps, but it too has nothing to do with sentence-like structures.

Moreover, the learning of conceptual frameworks and perceptual capacities appears to consist in the gradual adjustment of the overall configuration of the synaptic strengths and synaptic connections that join any neuronal population to any other. Those adjustments appear to be effected by a microlevel process called Hebbian learning, a process sensitive to the temporal coincidence of distinct excitations arriving at the synapses onto a common target neuron. Once again, propositional attitudes seem entirely absent from the process.

In all, concludes the eliminative materialist, the portrait of cognitive activity contemplated by modern neuroscience paints all terrestrial cognitive creatures, humans included, as vector coders and vector/matrix processors rather than as sentence coders and inference-drawing processors. Sentential processing, even for us, would appear to be a deeply secondary or tertiary cognitive skill. This conclusion is consistent with the very high levels of cognitive skill retained by humans who suffer left-brain global aphasia, a familiar malady in which the parts of the brain specifically devoted to language processing (roughly, Broca’s area and Wernicke’s area) are selectively destroyed by stroke. These people cannot process or manipulate linguistic structures (i.e., propositional attitudes) at all, but their overall cognition is little impaired.

A third area of worry for the propositional attitude core of folk psychology concerns its broad range of explanatory failures. We spend at least a third of our lives asleep, but folk psychology has no account of what sleep is or why all cognitive creatures should need it. The nature of mental illness, in its many forms, remains a mystery from within the kinematics and dynamics of folk psychology. Things as diverse as autism and schizophrenia receive no worthwhile explanation from within folk psychology: not of their causes, nor of their natures, nor of their cures. Sigmund Freud famously attempted a story of unconscious propositional attitudes to address the wide-ranging phenomena of mental illness, but that story failed utterly to explain the two severe maladies just mentioned, it was based on no systematic experimental testing, and it has since proved to be a therapeutic bust.

Our capacity for creative imagination and the basis of intelligence differences between individuals are both utterly mysterious from the perspective of folk psychology. And these two features—similar to sleep, mental illness, learning, and memory—are hardly peripheral cognitive phenomena. They are
core dimensions of cognition. The failures here are not peripheral; they are central.

A fourth worry arises from the massive performance failures of the research program of classical artificial intelligence (AI), a research program that aimed to re-create human and animal cognitive capacities by programming digital computers to manipulate various sentence-like states according to structure sensitive rules. These digital/serial machines were deliberately built so as to realize precisely the kind of cognitive kinematics and dynamics portrayed in folk psychology. Being electronic, they also had a millionfold speed advantage over biological brains. Yet they proved to be surprisingly inept and disappointingly slow at most of the cognitive skills displayed by biological brains. If we construe that failed research program as a test of the hypothesis that high-level cognition consists in the manipulation of sentence-like states according to structure-sensitive inference rules, then we are here looking at a presumptive refutation of that hypothesis. Thought, it would seem, doesn’t work that way.

A fifth and final worry is based on a historical induction. So many of humankind’s past theories and conceptions of natural phenomena have proved to be completely cockeyed: Think of the stories embraced in antiquity concerning the nature and causes of motion, the structure and makeup of the heavens, the nature of life, the nature and causes of disease, the age and origins of the Earth, the behavior of the oceans, and the causes of storms. Given that we were so utterly wrong about all these prominent phenomena, runs the argument, why expect that we would correctly apprehend the essential activity of the biological brain, something far more complicated than any of the phenomena just listed? Folk psychology, alone among the folk theories of antiquity, survives unchanged to this day and still constitutes our bedrock conception of what cognition consists in. But what are the chances that we got that one right, when we got all the other ones so wrong?

**Some Criticisms of Eliminative Materialism**

A common response to the preceding considerations is simply to resist the assimilation of our folk psychological conceptual framework to a theory and thereby to free it from the evaluative demands that an acceptable theory must meet. After all, if it isn’t a theory to begin with, it needn’t be threatened by the sorts of empirical failures cited above, even if they are real. This popular response plays down the predictive and explanatory functions of our folk framework for mental states and plausibly emphasizes its many social, normative, and practical functions instead.

A further, and independent, challenge to eliminative materialism concerns the identity and character of the millennial conceptual framework that is expected to supplant our current framework in all its many functions. If our current self-conception is an outright mistake (we have a right to ask), what alternative conception might actually replace it? Here the eliminative materialist must concede at least a temporary embarrassment, for the several cognitive neurosciences are still a long way from completing a theory that might be equal to this genuinely demanding task.

In purely scientific contexts, the vector-coding, matrix-processing story sketched above may well mature into a general theory of the real nature of all biological cognition. It paints a very different picture of the basic nature of representation and computation within the brain, and it might help us reconceive the many other dimensions of cognitive activity as well, much to our explanatory, predictive, normative, and manipulative advantage. A parallel result, note well, has already happened in the domain of our hidden metabolic and biochemical activities and in our scientific grasp of and control over the avowedly normative matter of our health. The four humors, the *élan vital*, and the sundry malevolent spirits of medieval medicine have been entirely eliminated from our modern medical ontology. We now speak regularly of bacteria, viruses, cancers, and other microbiological and biochemical matters.

In social contexts, however, it is much harder to predict what will happen. Conceivably we might someday learn to speak systematically of population-coding vectors and vectorial transformations within our brains, even at the dinner table and in the marketplace. But this is not immediately plausible. On the other hand, humankind has embraced new conceptual frameworks before, sometimes eagerly. For example, the Freudian psychoanalytic framework invaded the conversational vocabulary of 20th-century Europe and America, even though it lacked any genuine scientific credentials. The eliminative materialist leaves off by urging us, if only
in hope, to consider what a framework with real muscle might do. Reflect once more on the example of modern medicine.

Paul M. Churchland

See also Folk Psychology; Idealism; Mind-Body Problem; Realism and Instrumentalism; Reductive Physicalism

Further Readings


EMBARRASSMENT

Embarrassment is an emotion that arises in social contexts or when thinking of social situations. The experience of embarrassment is subjectively unpleasant and is accompanied by distinct physiological and behavioral changes. In some situations, the desire to avoid embarrassment can lead to negative personal and social consequences. However, embarrassment can also serve important beneficial social functions and may have evolved to do so.

Accounts of Embarrassment

There are two prominent accounts of what triggers embarrassment. According to the social evaluation model, embarrassment is elicited by the anticipation of negative evaluation by others. People become embarrassed when they perceive that the social image they want to project has been undermined and that others are forming negative impressions of them. However, this model does not readily account for all cases of embarrassment. For example, embarrassment can occur in situations that are entirely positive and that do not reflect negatively on the self in any way, such as when friends sing “Happy Birthday to You.”

According to another view, the awkward interaction or dramaturgic account, embarrassment arises when one anticipates the disruption of smooth social interaction or when one is unclear about the social expectations governing behavior. Based on this theory, embarrassment is not due to a person worrying about making a bad impression per se but rather is due to not knowing how to behave. Research suggests that a single theory probably is not adequate to account for all incidents of embarrassment and that both of these accounts have merit.

Development of Embarrassment

Embarrassment develops later than many emotions (such as anger, fear, and jealousy), usually emerging between 15 and 24 months of life. It appears to require that a child has some clear knowledge of the self (e.g., shows the ability to recognize that the reflection in the mirror is of the self). The importance of the awareness of the self is inherent in the common description of embarrassment as a self-conscious emotion.

Functions of Embarrassment

Relative to living alone, group living provides many advantages but also has its challenges. According to several theorists, embarrassment evolved as an instrument to help undo or prevent negative social consequences in situations where a person has unintentionally violated a social norm. The basic premise is that those who experienced and expressed distress over concerns with others’ impressions of them were more likely to survive as reproductive members of the group than those who acted with disregard for others’ opinions. Disregard for others’ reactions might have led one to be ostracized from the group or perhaps even killed.

Embarrassment likely has three basic functions. First, it serves as an appeasement gesture to others by signaling that the violation was unintended and that it will not likely occur again. Second, the intense dread of experiencing this emotion likely deters one from repeating whatever behavior triggered the emotion. Third, embarrassment motivates an individual to undo the social damage and restore the regard of others.

Findings from several studies are consistent with such functional accounts. For example, displaying embarrassment after an accidental mishap can lead to one being liked more and to children being punished less by their mothers, which provides some support for the theory that embarrassment is akin to an appeasement gesture, triggering a positive
response in onlookers. Research has also found that embarrassment can motivate individuals to engage in more prosocial behaviors such as volunteering. Other research suggests that expressions of embarrassment may have a cathartic effect. For example, people who were given the opportunity to show their embarrassment got over it more quickly than those who were not given such an opportunity. Thus, although people dislike showing embarrassment, it may be that once embarrassment is triggered, it is difficult to stop feeling uncomfortable until they behave in a functionally appropriate manner. Further evidence for this idea comes from work that finds that people are particularly likely to engage in apologies or acts of remediation after doing something embarrassing.

**Nonverbal Expressions and Physiology**

Embarrassment is accompanied by somewhat distinct nonverbal expressions and physiological indices that unfold in a unique fashion over time. The embarrassment expression often includes touching one's face, frequent gaze shifts, looking down, smiling, and attempts to control or inhibit the smile (e.g., by pressing the lips together). The smiling that occurs during embarrassment appears different from that of amusement. During amusement, two muscles are simultaneously contracted: the zygomatic major, which pulls the corners of the mouth up, and the orbicularis oculi, which crinkles the skin around the eyes (often called the Duchenne smile). In contrast, during embarrassment, the lips turn up without the crinkling of the eyes. Blushing frequently occurs during embarrassment, starting with a sharp increase of blood flow that is followed by a slower rise in facial temperature. The blood flow increase is responsible for the appearance of the blush. However, one's experience of blushing is based on perceiving temperature change. Hence, blushing is likely to be seen by others well before the embarrassed person feels it. Heart rate and blood pressure both initially increase, but then heart rate tends to drop while blood pressure continues to rise. This pattern of changes is different from what is seen in some other emotions and may be unique to embarrassment.

**Negative Consequences**

The desire to avoid embarrassment can lead people to put their best interests as well of that of others at risk. For example, fear of embarrassment may be involved in the frequent failure of onlookers to intervene during potential emergency situations. The bystander fears looking foolish if he or she responds as if the situation is a crisis and then turns out to be wrong. Failing to take action in such situations is irrational. The price paid for not taking action in an actual emergency far outweighs the minor cost of looking silly if the situation is not an emergency. Yet people often wait to respond.

Fear of embarrassment also contributes to unsafe sexual practices, such as failure to obtain and use condoms, and dissuades many people from getting medically recommended exams such as pap smears, mammograms, and colonoscopies. Some recent research suggests embarrassment serves as an obstacle in medical situations far broader than just awkward medical exams. Despite being worried about a potentially serious symptom, one third of an adult sample reported that they had failed or delayed to seek medical care to avoid looking silly if the symptom turned out to have a trivial cause. The symptoms that were most frequently mentioned in these cases were those that might indicate cardiac distress (e.g., chest pain). Delaying medical care in such circumstances is especially unwise since the negative effects of heart attacks can be greatly reduced by thrombolytic agents and emergency angioplasty. However, such procedures require rapid treatment to be effective. Thus, it appears likely that delaying or not seeking medical attention because of fear of embarrassment may be a substantial cause of avoidable morbidity and mortality.

*Christine R. Harris*

**See also** Emotion, Cultural Perspectives; Emotional Recognition, Neuropsychology of; Facial Expressions, Emotional; Guilt

**Further Readings**


**EMERGENCE**

The term *emergence* is used in a variety of (often incompatible) ways in the philosophic and scientific literature. However, *emergentism* is always a thesis about properties, and most versions share the following components: Emergent properties are (i) dependent upon, (ii) determined by, and (iii) not deducible from, basic physics. A highly diverse range of properties has been held, at various times, and by various philosophers and scientists, to be emergent. Examples include mental properties, most often consciousness; chemical bonding; and ordered patterns such as tornados in chaotic systems such as the weather. To say of such properties that they are emergent is to say something about the way they are related to the particles, properties, and laws of physics. This entry first clarifies the related notions of emergence and nondeducibility, then distinguishes three distinct forms of emergence with respect to the basic physical level. Let us begin with clarification of the key notions.

Emergent properties are

1. instantiated only by complex physical particulars,
2. determined by structural physical properties of their bearers, and
3. nondeducible from physics.

Emergentists about the mind, for instance, think that at least some mental properties (e.g., beliefs, desires, sensory experiences) satisfy (i–iii). Focus on this kind of emergence. Component (i) contrasts emergentism with Cartesian substance dualism—emergentists deny any nonphysical mental substance in which our mental lives take place. Mental properties are possessed only by things that are fully composed of physical parts (e.g., brains). Component (ii) holds that instantiation of certain structural physical properties is sufficient for instantiation of emergent properties and entails that you and I cannot differ mentally unless we also differ physically. Components (i) and (ii) are common to most extant theories of mind, emergentist or not, and are often jointly referred to as the supervenience of mental properties on the physical. What distinguishes emergentism from these other positions is component (iii). Nondeducibility makes emergent properties surprising, novel, and unexpected with respect to the physical properties they emerge from—no scientist could deduce your mental state merely from physical knowledge about your brain, body, environment, and so on. The broad consensus on defining emergence in terms of nondeducibility hides significant differences. Just as there are several ways in which you can fail to have blue eyes, so there are several ways in which a property can fail to be deducible from physics.

**Deducibility**

Arguably the only way to deduce a property from physical properties and laws is to complete a functional reduction of the property to be deduced. Understand a functional property to be a property defined by its causes and effects. Being a mousetrap is a functional property shared by many physically diverse machines—mousetraps differ in physical constitution, and they trap mice in a range of different ways. All that those machines have in common is that when you input a live mouse, they output a dead (or in some cases live but captive) mouse. The property of being a mousetrap is what all mousetraps have in common, and so is defined by what mousetraps do—by the way they function. Suppose we are trying to deduce a Property Q from physical properties, entities, and laws. According to Jaegwon Kim’s theory of functional reduction, we must

a. work Q into shape for deduction by conceiving it as a functional property (i.e., Q = the property of being so structured as to play Causal Role R); and

b. find a physical Property P that plays Causal Role R and explain how it does so in terms of basic physical laws and properties.

If we can complete these two steps, we are in a position to deduce which things have Property Q. An example will help. Water (H₂O) is highly cohesive compared to chemically similar compounds such as hydrogen sulfide (H₂S). For instance, H₂O boils at 100 Celsius (C), but H₂S boils at −60.266°C; so H₂S is gaseous at room temperature, whereas H₂O is liquid. Let Property Q = the cohesiveness of water, and suppose we want to deduce Q from basic physical laws and properties alone. First, we must understand cohesiveness in functional terms; let’s say
Emergence cohesiveness equals the property of having mutually attractive parts. Now if we can use physics to predict whether and to what extent x’s parts attract each other, we will thereby have deduced x’s cohesiveness. We know that water has \( \text{H}_2\text{O} \) molecules as parts. To deduce the cohesiveness of water from physics, we need to show that \( \text{H}_2\text{O} \) molecules attract each other to a degree that explains (for instance) why water is a liquid at room temperature. It turns out that we can do exactly this. Owing to their shape, \( \text{H}_2\text{O} \) molecules are electrostatic dipoles. Coulomb’s law tells us that \( \text{H}_2\text{O} \) molecules will exert an attractive force on each other; this is known as **hydrogen bonding**. \( \text{H}_2\text{S} \) molecules do not form hydrogen bonds, and so hydrogen sulfide is less cohesive than water. The deduction outlined above is a *reduction* because it shows us that the cohesiveness of water is *nothing over and above* the properties and relations of its molecules.

There are several ways for a property to be nondeducible, corresponding to the ways in which functional reductions can fail. Deducibility fails for a Property P if

1. P is not exhausted by its functional role, *or*
2. P has a causal role that is not occupied by a physical property, *or*
3. neither (1) nor (2) but *we cannot complete the deduction of P from physics because of our own conceptual and/or computational limitations.*

There may be other forms of nondeducibility, but conditions (1), (2), and (3) are at least independently *sufficient* for nondeducibility. Condition (1) will be met by properties with fundamentally qualitative natures (whether or not they also have functional roles). Such properties cannot be deduced by functional reduction because Step (a) of the reduction cannot be completed. Condition (2) will be met by any property that has novel, irreducible causal powers with respect to the physical. Such properties cannot be deduced because Step (b) cannot be completed. Conditions (1) and (2) are each sufficient for nondeducibility in *principle*; condition (3), on the other hand, will be met by any property that is in principle deducible (a property that meets neither of the two other conditions) but is *in practice* nondeducible due to certain facts about ourselves, which facts typically vary from case to case.

Given the preliminary characterization of emergence given in the introduction to this article, there will be as many kinds of emergence as there are ways for deducibility to fail. We can usefully divide them, however, into two categories: **ontological** and **epistemological**. Ontology is the study of what there is; correspondingly, ontological emergentists think that emergent properties are nondeducible because they are *something over and above* the physical. A supervenient property meeting either of conditions (1) or (2) will be ontologically emergent. Properties that meet (1) need not meet (2), and vice versa. Furthermore, (1) and (2) are compatible in that there might be a property that met both conditions, so counting as doubly nondeducible. In this entry, the term **weak ontological emergence** is used for properties that supervene on the physical and meet (1) but not (2), and **strong ontological emergence** for supervenient properties that meet (2), regardless of whether they meet (1). Epistemology is the study of knowledge; correspondingly, epistemological emergentists think that emergent properties are nondeducible because of limitations in our cognitive and/or conceptual abilities. A supervenient property meeting condition (3) will be epistemologically emergent. Certain properties that are nothing over and above the physical might be nondeducible for a range of reasons. For instance, our way of thinking about a property might make it very difficult for us to functionally conceive it, or the complexity of the physics involved might make it in practice impossible for us to deduce it. We may now appeal to the three different ways in which deducibility can fail in order to more precisely characterize three distinct kinds of emergence.

**Three Kinds of Emergence**

We will first give a broad characterization of the two kinds of ontological emergence, and through this, characterize epistemological emergence. Think back to the case of the cohesiveness of water. Cohesiveness is (a) exhausted by its causal role, and (b) *physically realized*, in that physical properties play the causal role that defines it. It is these facts that enable us to deduce the cohesiveness of water from physics. The deducibility of cohesiveness makes it transparent how it is related to the physical properties of \( \text{H}_2\text{O} \) and to the physical laws that govern its behavior. Once those properties and laws are fixed, we
get cohesiveness *for free*, as it is fully realized by hydrogen bonding. Physical realization of a Property P thus explains *why* P supervenes on the physical. Things are different with ontologically emergent properties. Consider a Property P1 that meets condition (1) for nondeducibility. Let P1 be a qualitative property that is defined by its *qualitative* nature. P1 is not physically realized since it is not defined by a functional role. The supervenience of P1 on the physical therefore requires *bridge laws* connecting it to its physical base. Now consider a Property P2 that meets condition (2) for nondeducibility. P2 may well be functionally defined, but it has no physical realizer. We also need bridge laws to ground the supervenience of P2. Ontological emergentists (both weak and strong) thus hold, in addition to (i–iii) above, that emergent properties are

iv. not physically realized and connected to the physical by irreducible “bridge laws” that are not laws of physics.

Suppose mental properties to be ontologically emergent, and that God has created all the basic physical entities, and made the laws of physics true. His work is not yet complete: He must still make the bridge laws true, thereby making it so that some physical things have mental properties. Because there are bridge laws, true in the actual world, relating physical and mental properties, it is not possible in the actual world for two physically indistinguishable individuals to differ mentally. However, there are possible worlds that are complete physical duplicates of this one but at which the bridge laws do not hold and so that have no mental properties. This makes any kind of ontological emergence inconsistent with *physicalism*. Physicalists think that everything that exists is nothing over and above the physical, so they must hold that any possible world that is physically indistinguishable from our world is indistinguishable in all other respects as well. If any properties in our world are ontologically emergent, then physicalism is false.

Properties contribute causal powers to the things that possess them—for instance, the momentum of a brick, not its color, contributes to the brick the power to break windows. Components (i–iv) of emergence are common to both weak and strong emergence and so do not distinguish emergent properties that do, from those that do not, contribute *novel* causal powers to their bearers. Let us add a fifth component to the existing four. Emergent properties

v. contribute powers over and above those contributed by the physical properties they emerge from.

This contribution of additional powers is typically called *downward causation* and violates the *causal closure* of the physical domain (discussed further below). We may now more precisely characterize our three kinds of emergence:

Strong ontological emergence = all of (i–v)

Weak ontological emergence = (i–iv); not (v)

Epistemological emergence = (i–iii); neither (iv) nor (v)

Both strongly and weakly ontologically emergent properties violate physicalism and require bridge laws to connect them to the physical properties they supervene upon. Strongly emergent properties do, while weakly emergent properties do not, contribute causal powers over and above those of the physical properties they supervene on. Epistemologically emergent properties, by contrast, are nothing over and above the physical, do not require bridge laws, and are nondeducible solely for practical reasons. The remainder of this entry gives examples of properties that have at one time or another been taken to be emergent in each of the senses characterized above.

**Strong Ontological Emergence**

Suppose mental properties are strongly emergent. If we try to predict how brains behave merely by summing the causal contributions of individual neurons, then we will “miss out” on the extra powers contributed by the mental properties and get the predictions wrong. By contrast, if we can explain the behavior of the brain just by reference to physical laws and properties, then mental properties do not exert a downward causal influence. A common position in the philosophic literature is that science has already progressed far enough to show that there is no downward causation from anything
nonphysical to anything physical. Physics forms a closed, complete system such that everything that happens within the physical domain can be fully explained without reference to anything outside it. If physics is causally closed, no properties are strongly emergent. Even if there is now ample evidence to support closure, it was not always so. The physics of C. D. Broad’s day, for instance, could not explain all the functional characteristics of complex chemicals, hence Broad’s theory that those characteristics were strongly emergent. It seemed to Broad as though chemical compounds had functional properties that were not physically realized, but emerged according to irreducible bridge laws, and that made causal contributions in addition to those of their physical components.

By way of illustration, imagine that the cohesiveness of water is strongly emergent. An easy way to imagine this is to suppose that H₂O molecules are not dipolar, so that there is no hydrogen bonding in water, but that water is nonetheless highly cohesive—that is, is liquid at room temperature, has a higher boiling point than H₂S, and so forth. The attractive forces between molecules will not be due to physical force-generating properties (e.g., Coulomb forces due to charge) but rather due to the emergent cohesiveness of water. Suppose that we have physical theories that fully explain and predict the behavior of hydrogen and oxygen atoms in isolation. Those theories will predict that water is a gas at 20°C, just as our current theories tell us that water would be a gas at 20°C if its molecules were not dipolar. Note that we can still complete Step (a) of our deduction, for even if cohesiveness were strongly emergent, we could still conceive it functionally. What we could not do, if cohesiveness were strongly emergent, is find a physical realizer for it. The success of modern physics in reductively explaining the characteristics of complex chemicals is no small part of the reason why Broad’s emergentism is no longer taken seriously.

**Weak Ontological Emergence**

Arguably the most plausible candidates for weak ontological emergence are *phenomenal properties*. Consider Frank Jackson’s famous example of Mary the color scientist who learns all the physical facts about color perception while locked inside a black and white room, without *seeing* anything colored. Suppose Mary is very clever and also knows all the facts that are deducible from physics. Intuitively, when she first sees a red tomato, she learns something—what it is like (to borrow Thomas Nagel’s [1974] phrase) to see red. Such examples convince most philosophers that phenomenal properties are not deducible from physics. It is also widely agreed that such properties are nondeducible because it is very difficult (perhaps impossible) to *conceive* of phenomenal properties, such as the visual redness of a red tomato or the painfulness of a pain, in purely functional terms. Understanding pain as a state caused by tissue damage, which causes you to say “ouch,” and so on, seems to leave out the most important part—its painfulness. Emergentists such as David Chalmers say this is because phenomenal properties are importantly different from functional properties such as cohesiveness, in that what is essential to them is the way they feel and not the way they behave. The phenomenal emerges from the physical according to bridge laws but does not add anything to it causally. On this view, there is a possible world physically identical to this one throughout its history—in which a physical doppelgänger of you is reading this article—but in which nothing is conscious. Weak ontological emergentists think that emergent properties are nondeducible because they are not functional properties, and this is why we cannot conceive them as such.

**Epistemological Emergence**

Type-B *physicalists* (this term is from Chalmers) agree that phenomenal properties are not deducible from physics and that this is due to our inability to conceive them in functional terms, but they deny that phenomenal properties are ontologically distinct from the physical. On this view, painfulness is a functional property, similar to cohesiveness. Our reluctance to accept this is explained by reference to the special nature of phenomenal *concepts* rather than the *properties* those concepts represent. We make a mistake when we think consciousness cannot be physical—the kind of mistake Lois Lane makes when she thinks Clark Kent could not be Superman. Phenomenal concepts are distinctive ways of thinking about the brain, which fool us into thinking that functional analyses of consciousness leave something out. Type-B physicalists think that phenomenal properties are nondeducible because phenomenal
concepts make it very difficult for us to conceive of those properties in functional terms. The concept of pain, for instance, makes it very difficult for us to see pain as a property exhausted by its function. Being in pain, however, is as functional as being a mousetrap. It is our way of thinking about pain that makes it nondeducible, not the nature of pain itself. Hence, the form of emergence here is epistemological.

Type-B physicalism is not the only kind of epistemological emergence. Not all mental properties are difficult to conceive functionally—beliefs and desires, for instance, seem more amenable to such analyses than phenomenal properties. Still, one might say, even these properties are nondeducible in practice, because of the extreme complexity of the physics involved. The forms of emergence considered thus far are synchronic—phenomenal properties, for instance, are taken to be nondeducible from the physical properties brains have at the same time. Weather systems provide an example of diachronic epistemological emergence (as do other systems whose dynamics are chaotic). Since the weather at any given time is extremely sensitive to slight changes in earlier conditions (the butterfly effect), it is impossible for us to deduce exactly when, in the future, a weather pattern such as a tornado will appear. The property of being a tornado is functional, and tornadoes are nothing over and above the air molecules that compose them. Still, we cannot deduce where or when they will occur. Such patterns are therefore epistemologically emergent with respect to physical conditions at earlier times. In general, epistemological emergence about a domain of properties differs from ontological emergence in that it explains the nondeducibility of emergent properties by reference to human nature rather than to the nature of the emergent properties. In the scientific literature, emergence is almost always used in its epistemological sense, and care must be taken not to conflate this usage with the far more controversial ontological variants discussed above.

Conclusion

Progress in science has reduced the appeal of strong ontological emergence through mounting evidence that all causal powers that physical things possess are contributed solely by their physical properties according to physical laws. Weak ontological emergence about consciousness, however, is more popular, and lively philosophical debates over the relative merits of this position and Type-B physicalism are ongoing. Since the laws of physics come out the same either way, it is difficult to see how any amount of scientific progress could help decide this issue. By far the liveliest area of current research on emergence concerns the various epistemological kinds, since these are consistent with the widespread commitment to physicalism among both philosophers and scientists.

David Yates

See also Anomalous Monism; Explanatory Gap; Mental Causation; Mind-Body Problem; Physicalism; Reductive Physicalism

Further Readings


Emotion, Cerebral Lateralization

The right and left cerebral hemispheres play selective, yet complimentary, roles in the perception,
experience, and expression of emotion. The extent to which each hemisphere contributes to the various components of emotional processing is widely debated. Two models of lateralized emotional processing have been put forward to explain diverse findings in emotion research: the right hemisphere model and the valence model, the latter of which was further developed into the approach-withdrawal model.

**The Right Hemisphere Hypothesis**

The right hemisphere model posits that the right hemisphere plays a specialized role in the perception, identification, and expression of emotional stimuli. It dates back to observations made by neurologist Charles Mills as early as 1912 that brain-damaged patients with unilateral right-sided lesions demonstrate a decrease in emotional expression. The right hemisphere hypothesis was later elaborated by Guido Gainotti to explain the catastrophic reaction of patients with left-hemisphere (LH) lesions and the indifferent or euphoric reactions of patients who suffered right-hemisphere (RH) lesions. Gainotti believed that RH damage led to an inability to emotionally appreciate the damage that the patients had incurred, while LH lesions resulted in an emotionally intact understanding of the amount of damage—hence, a normal secondary catastrophic reaction to a catastrophic event.

Experimental evidence to support RH specialization for emotion perception and identification includes a left visual field advantage for emotional stimuli, in that stimuli presented to the left side of the visual field, which leads to predominantly RH processing, are responded to more quickly and accurately. Joan Borod demonstrated that patients with RH lesions are more likely to be impaired in identifying emotional stimuli such as faces, emotional tone of voice (prosody), and words or sentences with emotional content. In split-brain patients, individuals who have had the major pathway connecting their two hemispheres (the corpus callosum) severed, both hemispheres are found to be equally adept at recognizing facial emotions on standard tasks. However, there is a RH advantage for distinctly nonverbal emotion discrimination tasks, which suggests that RH specialization for perceptual discrimination of nonverbal stimuli may also contribute to emotion-processing advantages.

Regarding expression of emotion, the left side of the face has been found to be more dramatic and expressive than the right side across all major emotion categories (happy, sad, mad, angry, surprised, disgusted) although this finding is more robust for negative emotions. Patients with RH damage also demonstrate impairments in expressing emotional prosody in both spontaneous and posed (intentionally produced) conditions. Additionally, patients with RH lesions are impaired in their electrodermal response to emotional stimuli relative to LH damaged patients. Finally, greater right lateralized electrophysiological activity is found during self-reported recall of emotional reactions to emotionally evocative visual material and when individuals are asked to generate emotional imagery.

**The Valence Hypothesis and Approach/Withdrawal**

The valence hypothesis posits that hemispheric specialization in emotion processing depends on the valence (positive or negative) of the emotion. Positive emotions are associated with greater LH activity and negative emotions with greater RH activity. The valence model was developed to account for the finding that the expression of positive emotions is not as strongly right lateralized as that of negative emotions, as well as for evidence of an LH processing advantage when positive emotional states are elicited by video clips, facial expressions, or reward-punishment contingencies.

The Wada test provides further evidence that emotional valence may be a differentiating factor in selective hemisphere involvement. This test involves injection of sodium amobarbital into the carotid artery of either the left or right hemisphere, which partially disables the hemisphere for a short period of time so that the functions of the preserved hemisphere might be assessed. Patients have been found to react differentially depending on the hemisphere that is anesthetized. During RH injections, patients may more commonly display indifference and euphoria, while depressive, worried mood and/or crying is more common during LH injections.

The valence hypothesis was later elaborated to include approach or withdrawal related attitudes and behaviors as a factor that might differentiate selective hemisphere involvement. Approach behaviors are those that bring an individual closer to an object or...
Emotion, Cerebral Lateralization

Emotion, Cerebral Lateralization

goal, and withdrawal behaviors are those that move one further away. Marcel Kinsbourne and Brenda Bemporad describe approach behavior as the movement to “organize and exert active control,” not just movement toward a desired object. The motivation for approach might be to exert control over a negative situation. This suggests that a selective LH advantage for approach behavior may not always coincide with positive valence. For example, Eddie Harmon-Jones and John Allen demonstrated an association between trait anger (anger that is dispositional and personality based) and increased left-frontal activity as demonstrated by lower electroencephalographic (EEG) alpha power. This is explained by individual differences in expectancies of control over a situation. If an individual feels angry about an expected negative outcome yet also feels competently motivated to overcome the situation, then more approach behavior and LH activity is demonstrated.

Nathan Fox and Richard Davidson demonstrated an association between approach-related anger and LH activity in two EEG studies involving infants. The first study demonstrated that 10-month-old infants who reached for their mothers during a task showed more left-frontal activation than those who did not. Additionally, infants who cried in response to maternal separation showed an increase in right-frontal activation. The second study demonstrated that 10-month-old infants who experienced anger upon maternal separation without crying showed more left-frontal activation. When infants were restrained, those who had shown visible signs of anger without crying also demonstrated more left-frontal activation.

Similar findings have been demonstrated in a different paradigm with adults. Eddie Harmon-Jones studied 77 undergraduates who independently paid at least one third of their tuition. His group manipulated the participants’ state anger (situational anger) by playing them a bogus radio broadcast announcing that the university had just voted to increase their tuition by 10%. While being monitored with EEG, half of the group was told that they could petition the decision, while the other half was told that the decision was certain. Greater left-frontal activity was observed in the group that experienced perceived control over the situation than the group that did not.

These studies argue against simple valence-based emotion discrimination as influencing hemispheric dominance and provide evidence that perceived control is an important component of the emotional experience. Such findings are in line with theories of learned helplessness that posit an association between decreased experienced control over a negative situation and withdrawal, apathy, and depression. Indeed, decreased left-frontal excitability is associated with depression, and stimulation of the left-frontal region has antidepressant effects. In 2008, the Food and Drug Administration (FDA) approved the application of transcranial magnetic stimulation (TMS) to the left frontal region as a treatment for depression. This highlights the importance of emotion lateralization research in developing treatment options for emotional disorders. Furthermore, improved understanding of hemispheric asymmetries in emotional processing can assist families of patients in coping with changed attitudes and behaviors following a lateralized brain injury. For example, research has demonstrated a correlation between the side of a stroke and premorbid/postmorbid attitudes and behaviors, such as greater depression and apathy associated with LH stroke.

In sum, the behavioral study of hemispheric laterality in emotional processing has produced inconsistent and sometimes contrasting results. A clear RH advantage in the perception and identification of emotional stimuli has been demonstrated; however, there is ample research to suggest an LH advantage in the processing of positive or approach-related stimuli and behaviors. Reasons for the discrepancies in findings might include the difficulty of designing an experimental paradigm that is free of confounds such as visuospatial processing advantages in the RH. Additionally, RH advantages in the modulation of arousal have been demonstrated, which might bias the RH in the processing of high-arousal emotional stimuli. Further work is needed to clarify whether hemispheric processing advantages might be related to specific components of the emotional experience (perception, expression, and experience), stimulus factors (verbal/nonverbal stimuli, low/high arousal), or situational factors (perceived control versus helplessness).

Karen Blackmon

See also

Emotional Recognition, Neuropsychology of; Facial Expressions, Emotional

Further Readings


**EMOTION, CULTURAL PERSPECTIVES**

Do people in different cultures experience and express emotions similarly? This is no less than one of the fundamental questions in understanding the nature of emotions. Early work has demonstrated that people around the world largely construct their emotional “diets” from similar ingredients, such as reacting with rapid heartbeat to danger or wrinkling their noses when noticing spoiled food. However, similarity in the basic ingredients does not preclude diversity. Using the same staples, human cultures produce marvelously diverse cuisines. Even dishes that are clearly similar have different flavors in different cultures. To what extent do cultural practices, norms, and values affect, evoke, reinforce, and/or modify emotional tendencies to produce culturally similar or diverse emotional experiences and behavior? Before describing the results of the studies that have begun to answer this question, this entry will briefly define the constructs of emotion and culture.

**Culture and Emotion: Definitions**

Emotions are responses to personally relevant events in one’s environment or internal representations of these events that involve loosely coordinated shifts in attention, subjective experience, behavior, thoughts, and bodily states. Emotions have the capacity to interrupt ongoing processes and facilitate an individual’s ability to flexibly and effectively respond to emerging threats and rewards. Emotions also involve mapping of these responses onto culturally meaningful emotional categories, such as happiness or angst. When emotional responses tap into widely shared and salient concerns, they become encoded in language and represented as emotional concepts.

Cultures are systems of widely shared goals, beliefs, norms, and values transmitted across generations and embodied in artifacts and behavior. These systems can affect emotions and in turn be affected by them. Culture can exert its influence on individuals’ emotions by fostering norms and ideas about whether or not particular emotions are desirable (e.g., a belief that interpersonal harmony can be threatened by expressions of happiness) and by what features of emotions are important (e.g., a belief that emotions affect the body). Another mechanism of cultural influence is repeated exposure to the behavior of others. For instance, although Paul Ekman’s seminal research showed that interpretations of intense facial expressions of emotions are remarkably similar around the world, recent studies show that in their daily lives individuals from different cultures show dialect-like differences in how they express their emotions. These differences are due to lifelong exposure to emotional expressions in our cultures. As a result, it is easier for us to correctly interpret emotional expressions by people from our own culture than those of others.

Because cultures are complex systems, in some cases emotions are shaped by the interaction between multiple aspects of culture. For instance,
although emotional expression is valued in Russian culture, it may be inappropriate for particular actors (e.g., men) or in particular circumstances (e.g., with strangers). Individuals rely on all relevant cultural knowledge to categorize changes in their experience and behavior as instances of particular emotions and to recruit information about these emotions. Depending on the nature of the recruited information (e.g., Is this emotion desirable? Can it harm one's well-being or one's relationships?), an individual may feel compelled to regulate the experience or expression of the emotion. Of course, people vary in their responses to cultural norms and in their ability and/or willingness to alter their emotions. It is, therefore, not surprising that significant within-cultural variability is observed in experienced and expressed emotions, even in response to standardized emotional tasks.

To return to our food metaphor, cultures differ not only in their favorite emotional dishes (or the extent to which emotions are considered valuable and functional in a given culture) but also in default “recipes” for emotions (or the extent to which components such as the presence of others are thought to be characteristic of emotions).

Culture and Ideas About Desirable Emotions

Norms regarding expression and experience of emotions and ideas about ideal or desirable emotional states differ across cultures. Cultures foster emotions that are conducive to culturally valued goals and discourage emotions that threaten these goals. For example, Western European and North American cultures foster an independent model of the self that foregrounds the autonomy and privacy of the individual. These cultures encourage experience and expression of emotions as a way of asserting one’s individuality. Although these norms apply to both positive and negative emotions, the experience and expression of positive emotions such as excitement and pride are particularly valued. In contrast, East Asian cultures foster an interdependent model of the self that foregrounds interpersonal ties. These cultures encourage moderation of emotions as a way to preserve relational harmony.

Even young children are encouraged to express (with an implicit goal of facilitating subjective experience) culturally appropriate emotions. Let us consider the example of pride. Because a feeling of pride in personal accomplishment and its expression reflects valued personal competence in European American culture, European American parents often encourage their children to feel proud. They may model celebratory high fives when a child excels (or even merely participates) in a competition or may state that they feel proud of him or her. In contrast, a feeling of pride under similar circumstances can threaten interpersonal harmony in East Asian cultures. Thus, Korean or Japanese parents may warn their children that it is important to be humble and mindful of the feelings of others and discourage excessive expressions of pride. Over time, these children are likely not only to adopt their parents’ values regarding pride but to also learn to moderate or enhance their feelings in accordance with these values.

Culture and Emotional Responses

Indeed, culturally desirable emotions, particularly positive emotions, are experienced and expressed more often. Individuals who live in cultural contexts that value expression of positive emotions such as pride (e.g., United States) feel and show these emotions more often and more intensely than individuals who live in cultural contexts that value moderation of these emotions (e.g., Japan). Although the association between cultural values regarding emotion and individuals’ experience of emotion is weaker for negative than for positive emotions, culture also affects the former. For example, the self-conscious emotions of shame and embarrassment can threaten self-worth in independent cultural settings. The same emotions can serve as useful signals of relational status in interdependent cultural contexts. Accordingly, these emotions are experienced less often and less intensely in independent than in interdependent cultural settings. Over a lifetime of living in their cultural environments, people become remarkably skilled at regulating their emotions in accordance with cultural norms. For example, during an international competition, it took athletes from different countries only about a second to alter their initially similar expressions of powerful emotions to show expressions that differed by the culture of the athlete.

Culture does not affect all aspects of emotions equally. Because what people feel and show on their faces has greater implications for interpersonal
relationships and is under greater voluntary control than bodily responses, cultural norms tend to shape the experience and expression of emotions. In contrast, cultural groups tend to show similar patterns of physiological responses to emotional stimuli. Moreover, not all reports of subjective experiences are similarly affected by culture. Cultural norms are more likely to shape global, retrospective, or hypothetical reports of emotion than to shape instantaneous reports, because the former depend on individuals’ cultural concepts of emotions more than the latter.

In summary, cultural norms regarding emotions are reflected in lived experiences of emotions, affecting the likelihood that an individual recognizes an instance of emotion, reacts to it, experiences it intensely, and displays it for others to see.

Culture and Conceptions of Emotions

Across cultures, emotions affect both individual and interpersonal concerns, regarding both the body and the mind. Cultural context can emphasize some of these emotional “ingredients” and de-emphasize others, thus affecting individuals’ experience and expression of emotions.

Socially Embedded Emotions

Emotions express and reproduce the culturally normative models of the self and social relationships. Although individuals across cultures recognize that emotions can play out in the privacy of one’s mind or in heated interactions with others, emotions are more likely to be construed as personal phenomena in independent cultures and as interpersonal phenomena in interdependent cultures. Individuals from interdependent cultures are more likely to describe emotions as interpersonal in nature, characterize them by interpersonal situations and their effect on relationships, use more relationships words when describing emotional episodes, take interpersonal context into account when interpreting emotional expressions, experience emotions in response to interpersonal triggers, share them with others, and feel them intensely for close others than are individuals from independent cultures. For instance, the mere thought of a family member intensifies the ability to experience and express emotions in response to a film for Asian American but not for European American individuals.

Somatically Embedded Emotions

Emotions are also both visceral and psychological. One can experience feelings of anger or love as powerful bodily sensations and as psychological states. Although emotions are embodied across cultures, some cultures emphasize somatic or bodily aspects of emotions. This tendency is expressed in emotional language and in concepts of emotions. For example, Ghanaians and Chinese ubiquitously refer to bodily signs and sensations such as red eyes or a burning sensation in one’s stomach when describing emotional concepts. Cultural emphasis on somatic aspects of emotion fosters awareness of one’s body during emotional episodes. It remains to be seen whether these differences translate into differences in the ability to accurately detect and regulate physiological changes.

Relationship Between Pleasant and Unpleasant Emotions

Another example of cultural differences in conceptions of emotions is the relationship between seemingly incompatible emotions. In some cultures, such as European American culture, feelings such as happiness and sadness are considered to be incompatible. On the other hand, East Asian traditional dialectic philosophies view pleasant and unpleasant feelings as compatible. As a result, individuals from East Asian cultures are more likely to co-experience opposing feelings than individuals from North American cultures. In bicultural individuals, the ability to tolerate opposing feelings is associated with engagement with East Asian cultures.

Conclusion

Although some “basic ingredients” of emotions (e.g., differential responses to threats and rewards or facial expressions used to signal emotional states to others) are similar across cultures, culture does affect individuals’ emotions. Individuals around the world are prepared to keenly attend to the features of their cultural environment and adjust to them when experiencing and expressing emotions and thinking of them. Cultural ideas about emotions that are considered to be ideal or functional can affect
the likelihood that individuals will regulate their emotional experience and expression by seeking out situations that maximize these emotions, cultivating attention to their elicitors, or expressing them more intensely. Cultural conceptions of emotions as interpersonal or somatic phenomena affect the likelihood that individuals will rely on these characteristics in their recognition of emotional episodes. In sum, because emotions and culture are intimately tied, the study of emotions needs to take cultural contexts into account when developing and testing new theories.

Yulia Chentsova Dutton

See also Emotion, Psychophysiology of; Emotion Regulation; Emotional Recognition, Neuropsychology of; Emotions and Consciousness; Facial Expressions, Emotional

Further Readings


**Emotion, Psychophysiology of**

Psychophysiological mechanisms involved in the processing of emotion have been most extensively studied in relation to aversive, threatening, or fear-related provocation. Beginning with the central nervous system circuitry subserving the processing of aversive or fear-related stimuli and then progressing to the several physiologic changes resulting from this mechanism, the following review will detail how the central nervous system affects change in psychophysiological systems, as well as the possibility that psychophysiological responses, in turn, guide and inform higher order cognition.

**Classical Fear Conditioning**

To study central and peripheral nervous system responses to emotional provocation, numerous tasks have been developed to experimentally elicit, albeit in likely attenuated form, emotional states and responses. Perhaps the most well-studied emotional provocation paradigm is that of classical fear conditioning, in which a previously neutral stimulus (e.g., a visually presented color square) is presented with or shortly before an aversive stimulus (e.g., an electrical shock). Over time, neutral stimuli paired with the aversive stimulus come to represent and elicit similar physiological responses to that of the aversive stimulus alone, while neutral stimuli unpaired with the aversive stimulus exhibit no such association either cognitively or physiologically. The aversive stimulus is typically referred to as the unconditioned stimulus (US) and the neutral stimulus is referred to as the conditioned stimulus (CS). Typical classical fear conditioning paradigms are used to study how fear associations are learned by the subject (acquisition) as well as how such associations diminish over time (extinction). An acquisition phase presents paired CS+ and aversive US stimuli together, as well as CS– stimuli with no subsequent US. A subsequent extinction phase presents CS+ without any further reinforcement of the US, as well as CS– stimuli. By observing brain activity, in the case of in vivo neuroimaging, and peripheral physiologic responses, in the case of psychophysiological recordings, the acquisition of the contingent relationship of CS+ with US and subsequent extinction...
of this association when reinforcement is halted can be assessed. As such, classical fear conditioning presents the opportunity to assess brain and physiologic responses involved in aversive stimulation, the development of learned fear associations, and the reversal or extinction of learned fear associations. As a result, this paradigm has been used extensively to study normal functioning of underlying mechanisms recruited in fear processing as well as in psychiatric and neuropsychiatric samples ostensibly marked by disrupted or exaggerated fear response.

The Amygdala, Innate and Learned Fear

Utilizing classical fear conditioning as well as other experimental methods, several researchers have attempted to identify central nervous system mechanisms involved in the initiation and modulation of the fear response. Studies in nonhuman animals have increasingly identified the amygdalae, nuclei in the medial temporal lobes, as being central to the processing of threat and resulting fear response. Amygdala function occurs in the context of direct and indirect interconnections with other brain regions that subserve behavioral modulation and inhibition (prefrontal cortex), learning and memory (hippocampi), sensory processing (thalamus, sensory and polymodal cortex), and initiation of peripheral physiologic response (hypothalamus). Previous work by Joe LeDoux and others has established the central role of the amygdala in the processing of threat and in learned associations of stimuli that are associated with threat. The amygdala itself exhibits subregions that appear to be functionally dissociable. The lateral nucleus of the amygdala is a major afferent (input) nucleus that receives sensory information and projects to both basal and central nuclei of the amygdala. The basal nucleus of the amygdala projects to brain regions identified with movement and motor programming (striatum), potentially important in executing motor activity in response to threat, while the central nucleus projects to key hypothalamic and brain stem sites, which initiate and coordinate autonomic and endocrinological processes as well as species-specific physiological and behavioral routines in response to threat. Beyond innate threat processing, previous work indicates that learned contingencies between neutral and aversive stimuli are subserved by synaptic plasticity in an interneuronal network connecting the lateral and central nuclei of the amygdala. CS+ and US information converge in lateral amygdala and, as a result of their co-occurrence, alter synaptic connections (via Hebbian plasticity) such that presentation of the CS+ in isolation yields similar output from the central nucleus of the amygdala. Extinction, the process by which learned fear contingencies are suppressed when the US is no longer associated with the CS+, is subserved by prefrontal, and specifically medial prefrontal, interconnections with the central nucleus of the amygdala. Excitatory (glutamatergic) projections from medial prefrontal cortex projecting to inhibitory (γ-aminobutyric acid, [GABA]ergic) regions of the central nucleus of the amygdala appear to play a role in extinction by the suppression of response to the conditioned stimulus coordinated in the central nucleus.

The Hypothalamus: Neuroendocrine and Autonomic Control

One important aspect of the central nucleus of the amygdala lies in its connectivity, together with the prefrontal cortex and hippocampus, via the bed nucleus of the stria terminalis, to the hypothalamus. The hypothalamus projects to the pituitary, through which neuroendocrine response is coordinated (hypothalamic-pituitary adrenal axis [HPA-axis]), and to the medulla, which is implicated in control of the autonomic nervous system. Each system is important in coordinating the organism’s response to emotional provocation and threat. In regard to neuroendocrine coordination, in addition to oxytocin and vasopressin, adrenocorticotropic hormone (ACTH) is released into the bloodstream by the pituitary, stimulating the adrenal cortices, which in turn secrete cortisol, a stress hormone implicated in the body’s reaction to stress. In regard to medulla-mediated autonomic system modulation, hypothalamic projections to medulla in turn affect both parasympathetic and sympathetic nervous system activity, influencing homeostatic and fight-flight responses. The central nucleus of the amygdala exhibits a mostly excitatory role in hypothalamic function and is instrumental in ACTH secretion.
after stress. In contrast, the hippocampus exhibits a mostly inhibitory role in hypothalamic function via a feedback loop mediated by cortisol. The ventromedial prefrontal cortex also exhibits a modulatory role on the HPA-axis with lesions in this region being associated with altered stress responsivity and with activity in this region being negatively correlated with amygdala activity and predictive of daily cortisol secretion. Several parallel changes in the peripheral nervous system and organ systems occur following stressful provocation and resulting hypothalamic output. These changes can be understood as improving the organism’s ability to address or cope with an initial threat. Neuroendocrine related changes, here focusing on those mediated by cortisol, include suppression of the immune response, increased availability of glucose (energy), suppression of the reproductive system and growth, and cognitive changes, including enhanced attention and memory formation. Significantly, both the amygdala and hippocampus are rich in glucocorticoid receptors, likely partly mediating the emotional and memory effects in the context of stress. Autonomic sympathetic and parasympathetic systems serve to modulate pupil dilation and constriction, gut motility, digestion, vasoconstriction and dilation, sweat production, heart rate, and bronchial dilation and constriction among other homeostatic processes.

**Psychophysiological Measures**

These various physiological changes, predominantly initiated and modulated by sympathetic and parasympathetic systems, are at the center of the several methodologies that allow for psychophysiological analysis of emotional provocation. The key assertion that the use of such measures share is that changes in measured activity in any psychophysiological domain represent recruitment of central nervous system and autonomic system mechanisms. While several such psychophysiological indicators have been studied in both human and nonhuman animals, electrodermal activity (EDA), electrocardiography (ECG), and startle electromyography (EMG) are here the focus for their frequency of use in the literature. EDA, ECG, and startle EMG activity have been proposed as a subset of potentially several *somatic markers*, discussed below, that might guide decision making or inform the organism of relevant emotional states.

**Electrodermal Activity**

Electrodermal activity (EDA), the measure of sweat gland activity in the skin, has been used extensively to study emotional provocation. In human studies, two electrodes are attached, typically to the fingertips, and a low current is passed. Changes in resistance occur as sweat gland (eccrine) activity increases or decreases. EDA can be measured over longer periods of time (*tonic* responses—skin conductance level [SCL]) as well as in relation to individual stimuli over shorter periods of time (*phasic* responses—skin conductance response [SCR]). Although methods vary, a window of 0.5 to 4 seconds after stimulus presentation is often used to analyze SCRs resulting from a discrete stimulus; latency from onset of the stimulus, SCR amplitude, and time between SCR onset and peak are typically assessed. In the example of fear conditioning described earlier, SCRs are generally exhibited following presentation of the US (e.g., an electrical shock), and, over time, specifically to neutral stimuli predictive of the US (CS+). Over the course of the learning (acquisition) phase of the experiment, SCRs to both CS+ and US stimuli generally decrease, ostensibly as subjects become habituated to the aversive US. In the extinction phase of the experiment, during which the aversive US is no longer presented, SCRs to the CS+ decrease further (extinction), as their significance and association with the US diminishes. While EDA has been used extensively in fear conditioning studies, it is important to note that phasic SCR responses are also elicited by stimulus novelty (orienting), surprise, arousal, and significance. Control of EDA appears to be solely through the sympathetic system, with premotor cortex, amygdalar, hippocampal, hypothalamic, and brain stem sites all being identified as important contributors to tonic and phasic responses.

**Electrocardiography**

Electrocardiography (ECG), the measurement of cardiac activity, has most often been used in assessing autonomic activity over longer periods. Heart rate (HR) and HR variability (HRV) have been associated with autonomic nervous system activity through sympathetic adrenergic beta-1 receptors and vagally mediated parasympathetic cholinergic receptors, which control acceleration and deceleration of
heart rate respectively. In healthy subjects, stimulus input, such as passively viewing unpleasant or pleasant pictorial stimuli, has been shown to correspond with a deceleration of HR, while induction of mental imagery of arousing pleasant and unpleasant stimuli corresponds with an acceleration in HR. While HR measures are informative in regard to collective sympathetic and parasympathetic involvement in cardiac function, analysis of the variability of heart rate yields important information in regard to specific contributions of each. HRV can be assessed in both time and frequency domains, with variability in time assessed by measuring latency between successive beats and variability in frequency assessed by analyzing the relative power of frequency components of the ECG signal. In regard to frequency analysis, high-frequency changes are associated with parasympathetic modulation of the heart because of the relatively fast degradation of acetylcholine, whereas low-frequency changes are believed to be influenced by both sympathetic and parasympathetic systems. Importantly, respiration rate also influences HRV (respiratory sinus arrhythmia) with beat-to-beat intervals becoming shorter during inspiration and longer during expiration. Previous work in posttraumatic stress disorder, a condition marked by heightened sensitivity to stress provocation, suggests resting variability is marked by reduced high frequency and increased low-frequency components in the PTSD group interpreted as increased involvement of the sympathetic system and reduced modulation by the parasympathetic system.

**Electromyography**

Startle has been extensively studied in human and nonhuman animals and represents the mainly physical reaction to sudden, unanticipated stimuli. In such situations, facial muscles may contract, particularly around the eyes, together with muscle contraction in the trunk, arms, and legs. Electromyography (EMG), specifically used to assess the startle response, has been used to assess baseline levels of startle as well as experimental manipulation of startle under specific emotional conditions. Facial EMG measurement refers to a set of measures that are sensitive to changes in the surface (skin) electromagnetic field created by the convergence of muscle action potentials in a motor unit. The startle eyeblink response (SER), has been used extensively in the study of emotional processing, and specifically refers to the use of EMG to measure latency and magnitude of eyeblink response (at the orbicularis oculi) when the subject is acoustically startled (e.g., after presentation of a loud, white noise burst). Prior work demonstrates that the magnitude and latency of the startle response can be manipulated given certain emotional or motivational stimulus conditions. For instance, using the example of fear conditioning above, subjects demonstrate larger and faster startle responses during presentation of the CS+, once its association with the US shock has been established, than to the CS−, which is not previously associated with the US. Research on the initiation and modulation of the SER suggests the presence of at least two basic pathways, with startle initiation subserved by a pathway from the cochlear nucleus to pontine reticular formation to reflex effectors, while startle modulation is subserved by a connection between the central nucleus of the amygdala and the pontine reticular formation.

**Potential Influence of Psychophysiological Reactivity on Cognition**

These peripheral effects of threat provocation and autonomic nervous system activation have been variously implicated in providing a bodily feedback loop to the central nervous system. Changes in physiology or behavior attendant to specific emotional processes has been previously theorized to constitute the feeling and conscious experience of given emotional states (James-Lange theory of emotion), although this view has been widely disputed and is no longer accepted. More recently, Antonio Damasio and colleagues have proposed the somatic marker hypothesis, which posits that changes in bodily responses to emotional provocation help guide higher order decision making. This work has utilized findings from individuals with localized prefrontal lesions, in whom affective and social judgments are observed to be disrupted, putatively as a result of damage to the medial prefrontal cortex. Such individuals have been studied using the Iowa gambling task (IGT) in which subjects are given the opportunity to pick from four decks of cards: Two decks give high financial reward but higher financial loss over successive trials, while the other two decks give low financial reward but also lower financial loss over successive trials. Subjects realize a net financial gain if sampling
is predominantly from the low reward–lower risk decks and a net financial loss from the high reward–higher risk decks. Early in the task, normal subjects cannot consciously elaborate the contingency (they cannot report on or describe the chances in the two decks), so successful performance relies on intuition, or a gut feeling, rather than declarative knowledge about the contingencies. Interestingly, individuals with medial prefrontal cortex damage perform poorly on this task; they do not adjust their decisions as efficiently over learning trials to maximize their reward, and they exhibit diminished electrodermal reactivity when selecting from high-risk decks and in response to negative, punishing decisions. Damasio and colleagues understand this as a diminished ability in the patient group to link to and utilize psychophysiological states resulting from negative affective experiences in decision making. While the somatic marker hypothesis has been extensively studied, potential methodological issues and alternate interpretations of these results have been suggested. Still, this work has done much to flesh out the possible participation of a psychophysiological feedback loop to higher order cognitive processing and decision making.

James C. Root and Kane Pryor

See also Anxiety Disorders; Emotion and Psychopathology; Emotional Recognition, Neuropsychology of

Further Readings

EMOTION, STRUCTURAL APPROACHES

Since the time of James and Wundt scientists have debated about the most elemental or basic building blocks of emotional life. How many elements are there? What is their nature? How should they be referred to? These are referred to as questions about the structure of emotion, and this entry surveys the two main answers that have been offered by psychology (discrete and dimensional approaches to emotion structure). Such questions are important to resolve, because the answers will dictate which constructs will best support scientific induction and allow for the accumulation of knowledge about emotion.

Discrete Emotion Approaches to the Structure of Emotion

A discrete emotion approach to the structure of emotion argues that certain categories (e.g., those described by such English words as anger, sadness, fear, happiness, and disgust) form the most basic elements of emotional life. Discrete emotion approaches have been consistently criticized throughout the history of emotion research because instrument-based measures of the face, voice, body, and brain in humans (i.e., measures that do not require a human perceiver) as well as animal studies have not revealed the existence of discrete categories.
Emotion, Structural Approaches

in nature (for recent reviews, see Barrett and colleagues). Some studies of cardiovascular measurements, electromyographic activity of facial muscles, acoustical analyses of vocal cues, and blood flow changes within the brain do show consistent differences between discrete emotion categories, but the larger body of evidence disconfirms anger, sadness, fear, happiness, and disgust as fundamental mental elements. This is because there is tremendous variability in the patterns observed within a single category, as well as low consistency in the patterns observed across categories. Nonetheless, perceiver-based measurements (as when one person judges the face or actions of another) often give evidence of clear categories.

Dimensional Approaches to the Structure of Emotion

A dimensional approach to the structure of emotion assumes that the words anger, sadness, fear, and so on name folk categories that divide up the continuous and contextually sensitive range of mental events consisting of highly variable measurable outcomes (e.g., facial muscle movements, peripheral physiology, behavior). These events are constructed from more fundamental building blocks, with affect as one key element. In English, the word affect means to produce a change. To be affected by something is to be influenced by it. In science, and particularly in psychology, affect refers to a special kind of influence—something’s ability to influence your mind in a way that is linked to your body. Historically, affect referred to a simple feeling—to be affected is to feel something. In modern psychological usage, affect refers to the mental counterpart of internal bodily representations associated with emotions, actions that involve some degree of motivation, intensity, and force, or even personality dispositions.

Dimensional models describe affect as having the properties of valence and arousal (Russell and Barrett), valence and intensity (e.g., Lang), positive and negative activation (e.g., Watson and Tellegen), positive and negative affect (Cacioppo and colleagues), tense and energetic arousal (Thayer), or approach and withdrawal (e.g., Davidson). The properties of affect, no matter how they are characterized, can be summarized using the affective circumplex (Barrett and Bliss-Moreau) (see Figure 1). The circumplex has two parts: a circle, which depicts the similarity or relatedness between affective elements (be they words or faces or reports of emotional

Figure 1 The affective circumplex

Note: The affective circumplex is represented as a circle embedded in two-dimensional Euclidean space. The circle represents the ordinal ordering of affect and emotion categories relative to one another. The valence (hedonic) property of affect is represented on the horizontal axis. The arousal (activity) property is represented on the vertical axis.
experience), and a set of axes, which are the dimensions themselves and represent the psychological properties that quantify what is similar and different about the elements. Not all dimensional models of affect incorporate circumplex assumptions, although they are usually depicted in a circular space. And although there are debates over which set of dimensions best describes affective space, all can be incorporated into the same circular structure (Carroll and colleagues). Dimensional approaches to the structure of emotion have been criticized because they cannot explain how the instances referred to by the English words anger, fear, and disgust are different from one another (for example, these are all unpleasant, high arousal states).

To some extent, dimensional approaches to the structure of emotion have been misunderstood all the way back to Wundt. Most modern treatments of Wundt focus on his model of simple feelings—what we would now call momentary affective states—as having three independent qualities—pleasant/unpleasant, arousing/subduing, and strain/relaxation. Some combination of these properties can be found in what are often referred to as dimensional accounts of affect and emotion. According to Wundt, valence, arousal, and intensity are not basic ingredients of an affective state but rather descriptive features of a unified state that is (or that at one time was) grounded in the visceral, kinesthetic, proprioceptive, and neurochemical fluctuations that take place within the core of body. These three properties define a multidimensional affective space that people inhabit. Because it was so well developed, Wundt’s dimensional account of affect largely overshadowed his broader psychological constructionist approach to emotion (i.e., the approach that all mental events are constructed out of a common set of psychological ingredients; for a historical review of psychological construction models of emotion see Gendron and Barrett; for an example of a modern psychological construction approach, see Barrett or Russell).

According to Wundt, emotions are “psychical compounds” that are constituted of more basic psychological elements that are simple and irreducible in a psychological sense. One element was affect, and the other was ideation. According to Wundt, emotions begin with an “inceptive feeling” that is affective in nature. The inceptive feeling is caused either by external sensory stimulation (what Wundt calls “outer emotional stimulation”) or internal stimulation arising from association or apperception (i.e., perceiving new experience in relation to past experience—what Wundt refers to as “psychical”). Next, an ideational process distinguishes different emotional feelings from one another. Although Wundt did not provide a clear definition of what an ideational process is, he argued that ideas were the revival of previous experiences. Finally, there is a terminal feeling, which is basically a more diffuse affective state that remains after the more intense feelings have dissipated—similar to a mood state.

Many models of emotion that are called “dimensional” actually follow Wundt’s psychological constructionist approach, where emotions emerge from the interplay of more basic elements that are thought to be the fundamental ingredients of the mind.

In modern psychological constructionist approaches to emotion (e.g., Barrett; Russell), valence and arousal are descriptive, not causal, properties of affect. Although it is possible to focus on one property or the other, people cannot feel pleasant or unpleasant in a way that is isolated from their degree of arousal. Furthermore, similar to Wundt, these approaches assume that emotions are not different in kind from mental states from thoughts, memories, and beliefs. It is presumed that every moment of your waking life is to some degree pleasant or unpleasant with some degree of arousal so that affect is a basic property of consciousness (Duncan and Barrett).

Other dimensional approaches assume that the descriptive structure of affect should be isomorphic with its causal structure, however, so that the best affective dimensions are those that are most causally plausible (i.e., the dimensions should reflect the processes that cause affective states). Accordingly, it has been claimed that certain dimensions (positive and negative affect or activation, for example) are more biologically basic, should be the preferred when anchoring a dimensional model of emotion, and are themselves the ingredients that make up emotional life. This idea has been heavily debated, however (for a review, see Barrett and Bliss-Moreau).

Although various models disagree on the specifics, all dimensional approaches to emotion assume that affect is realized in a neural reference space in the front of the brain (see Barrett and Bliss-Moreau). This reference space includes brain areas that are traditionally considered to be “emotional,” such as the amygdala and ventral striatum, as well as
paralimbic portions of prefrontal cortex that until recently have been considered “cognitive” (cf. Duncan and Barrett). These areas, along with other (including insula, thalamus, hypothalamus, and autonomic control centers in the midbrain brain stem) constitute a large-scale workspace that integrates representations of sensory information from the world with their somatovisceral impact. This workspace not only participates in the creation of emotion but also infuses every mental state with affective content. When core affect is in the background of consciousness, it is perceived as a property of the world rather than as the person’s reaction to it. It is under these circumstances that scientists usually refer to affect as unconscious. We experience a world of facts rather than feelings, and affect gives us a sense of confidence in those facts. This is why a drink tastes delicious or is unappetizing, why we experience some people as nice and others as mean, and why some paintings are beautiful, whereas others are ugly. When core affect is in the foreground of consciousness, it is experienced as a personal reaction to the world: We like or dislike a drink, a person, or a painting.

Lisa Feldman Barrett

See also Emotion, Psychophysiology of; Facial Expressions, Emotional; Unconscious Emotions, Psychological Perspectives

Further Readings


EMOTION AND MORAL JUDGMENT

Some philosophers and cognitive scientists have claimed that emotions are essential to morality, while others deny this. This entry focuses on the relationship between emotion and moral judgment. First, moral judgments are distinguished from the psychological states that lead to moral behavior, then competing theories of moral judgment are examined, and finally, empirical evidence linking emotions and moral judgments is reviewed.

Moral Judgment and Moral Behavior

Before surveying theories and evidence, it is important to distinguish between moral judgment and moral behavior. Moral behavior includes any behavior that we would praise as morally good. Moral judgments are judgments that assess the moral significance of something—judgments about what is right or wrong, good or bad, moral or immoral. It should be clear from this distinction that there can be moral behaviors without moral judgments. For
example, a person might rush to save a drowning child without first reflecting on whether it is the right thing to do. Conversely, there can be moral judgments without moral behavior. A person might realize that it is wrong to cheat but give in to that temptation.

This entry will focus on moral judgments, not behaviors, but it is important to note that emotions have been implicated in both. Researchers who study altruistic behavior widely assume that we are impelled to help others at personal expense because of our emotions. The principal affective response implicated in altruistic motivation is empathy or empathetic concern, which involves either feeling a negative emotion akin to one you observe in another party or feeling bad about another person's suffering. The term sympathy is sometimes used for responses of the latter kind, where the felt emotion does not match the observed emotion. Daniel Batson has shown that when people experience this kind of emotional connection with another party, they do not simply act to alleviate their own discomfort but rather work to help the person in need. When this happens, a person can be said to have an altruistic motivation for the altruistic behavior, meaning that the person is not motivated by an explicit desire to advance self-interest. In this situation, the actor need not make a moral judgment that helping is good; it is enough to recognize that someone else is in harm's way.

Emotional responses may also play a role in so-called virtuous behavior, independent of moral judgments. Building on this idea, Aristotle developed an influential approach to ethics, which is now known as virtue theory, according to which the goal of morality should be the cultivation of good character traits rather than the internalization of moral rules. Good traits may include generosity, friendliness, courage, self-respect, among others. Aristotle and many of his followers think of these traits as involving emotion regulation. For example, courage may involve having enough fear to avoid being foolhardy but not so much as to be a coward. Thus, emotions are implicated in morality in this view, but virtue is presumed to be achieved without explicit moral judgments: For example, the virtuous person does not act courageously because she judges that courage is good but rather because she has trained her emotions to make practical decisions that are courageous when courage is called for.

Moral judgments are appraisals of moral value. We can appraise an action, a rule, a character trait, and so on. There is ample evidence, as we will see, that emotions arise in the context of moral judgments, but these emotions differ from the ones that typically underlie altruism and behavior that we deem virtuous. Altruism is often motivated by empathy; someone will save a drowning child out of concern for its welfare. But if you judge that this heroic act is good, the feeling associated with that judgment is not empathy but rather gratitude, admiration, or elevation. These are emotions of positive appraisal. Heroism is less common than misconduct, however, so most moral appraisals are negative. We condemn those who do wrong.

Emotions associated with condemnation vary as a function of the norm that has been violated and the identity of the culprit. Paul Rozin and his collaborators have shown that we feel angry about crimes that violate an individual's autonomy (harms or rights violations), we feel contempt for those who disrespect social hierarchies and public goods, and we feel disgust at acts deemed unnatural, such as sexual taboos or, in nonsecular societies, crimes that violate divine strictures against moral purity. These emotions are directed at other people who violate norms. When we violate norms ourselves, the emotions felt are usually guilt and shame. Guilt is usually construed as directed toward an action, and shame is directed toward the self.

There is a link between moral judgments and moral behavior. For example, bad feelings directed toward others can promote punishment or social ostracism, and bad feelings directed toward the self can motivate reparation or withdrawal. But one can also make a moral judgment without acting on these emotions, if other motivations are overriding. Thus, judgment is best defined as an attitude rather than a behavior.

Theories of the Relation Between Emotions and Moral Judgment

There are many theories of how emotions relate to moral judgment. Most of these were originally articulated by philosophers, but some new proposals are also emerging in light of empirical research.

Members of the Scottish Enlightenment developed a theory called sentimentalism in the 18th century. According to this theory, emotions are components
of moral judgments. To illustrate, consider the judgment that killing innocent people is wrong. For the British moralists, this would be defined as a feeling of disapprobation directed toward the idea of killing. For Francis Hutcheson such feelings were comparable to perceptions that pick up on real moral facts. For David Hume, there are no moral facts, but our feelings make it seem as though there were. Killing feels like it is intrinsically wrong, but the wrongness resides in us (a view sometimes called projectivism).

In the 20th century, expressivists adopted the sentimentalist theory, and they added a linguistic thesis according to which the sentences that assert moral judgments express feelings rather than asserting facts. A related view, the sensibility theory, says that moral facts are response-dependent properties. A response-dependent property is a property that depends on subjective responses. For example, to say something is funny is to imply that it causes amusement, and to say that it is delicious is to say that it causes gustatory pleasure. Sensibility theorists say that to call something wrong is to say that it causes (or merits) emotions of disapprobation, such as anger or guilt. Expressivism and the sensibility theory disagree about the semantics of moral judgments—whether they refer to response dependent properties or merely express feelings—but they agree that we need to classify it. Is it a case of intentional killing? Of dishonesty? Of discrimination? Of torture? Such questions often require extensive rational deliberation. Sentimentalism does not deny this. It is committed only to the thesis that once an action has been classified, emotions arise, and the presence of those emotions is necessary for judging that the action is wrong.

The sentimentalist position is opposed by those who think emotions are inessential accompaniments of moral judgments. Immanuel Kant argues that emotions often prompt us to make erroneous moral judgments. He said that, instead, we should base moral judgments on rational decision procedures. Kant’s most famous suggestion is that we can determine which actions are permissible by assessing which actions we could will as universal laws. Kant admits that this method has, at its foundation, a respect for human dignity, but he does not think emotions are essential components of moral judgments and he thinks that we should not make judgments based on gut feelings. The theory is normative—it tells us how we should make moral judgments—rather than descriptive, but it does entail that we can make moral judgments without emotions. Kant’s rival, John Stuart Mill, also emphasizes the possibility of making moral judgments on rational grounds. He says we should do that which maximizes utility (i.e., happiness)—something we can assess quantitatively without depending on any gut reactions. Both theories can be regarded as forms of moral rationalism, because they limit the role of emotions and emphasize reasoning and rational principles.

Within contemporary psychology, several theories try to limit the role of emotions as well. Elliot Turiel argues that moral rules are those that are universal, independent of authority, and more serious than mere social conventions. For him and other domain theorists, we assess whether something is morally wrong by determining whether anyone has been intentionally harmed, not by introspecting about our gut feelings. Marc Hauser argues that people assess whether something is wrong by unconsciously classifying actions in accordance with innately constrained moral rules that specify, for example, that intentionally killing someone to achieve a good end is worse than killing someone as a foreseen side effect while pursuing a good end. When we judge that something is wrong, emotions arise, but they are not the causes of our moral judgment.
Between the sentimentalists and their opponents lie dual-process theories. According to these, some moral judgments are based on emotions, and others are based on nonemotional processes. One variant of this view says that we use emotions as heuristics to decide if something is wrong but that they are dispensable.

In summary, emotions are sometimes seen as components of moral judgments, sometimes as causes, and sometimes as effects. They are regarded as necessary for normal moral judgments by sentimentalists, whereas opponents of sentimentalism and defenders of dual-process theories say that they are not necessary.

Evidence Relating Emotions to Moral Judgment

Recently, empirical evidence has accrued to establish that emotions play a role in morality, and it has been marshaled in support of one or another of the aforementioned theories.

In cognitive neuroscience, studies suggest that moral reflection causes increased activation in brain areas that have been independently associated with emotion. Jorge Moll has shown that emotions are active when making judgments of moral wrongness compared to factual wrongness and when we look at photographs of moral violations. Joshua Greene and his collaborators have shown that emotions are active when people consider moral dilemmas compared to factual dilemmas.

Such findings are often interpreted as evidence for sentimentalism, but they do not firmly establish whether emotions are causes or effects of moral judgments nor whether emotions are necessary. Greene found that emotions are much more active when considering dilemmas that involve physically harming someone to save lives, compared to dilemmas in which one causes harm to someone indirectly to save lives. For example, people show stronger emotions when imagining pushing someone in front of a runaway trolley to save five people who are standing on the track, compared to imagining diverting the trolley to an alternate track, where it will kill one instead of five. Greene uses this to support a dual-process theory on which emotions are used to make decisions about physical battery, but not in cases where we can assess moral behavior by just looking at the total outcomes of action and inaction. The data are consistent with this interpretation but are equivocal since emotions are active in all the moral dilemmas studied, just to varying degrees.

To assess whether emotions are causally implicated in moral judgments, psychologists turn to behavioral methods. Much of this research focuses on disgust. Simone Schnall and her colleagues have shown that disgust induction (though filth, disgusting movies, autobiographical recall, and foul smells) can increase the severity of moral judgments. Similar results have been obtained though hypnotic induction of disgust, and people who are disposed to feel more disgust than others tend to judge mores related to sexuality more harshly. Jonathan Haidt has shown that disgust can bolster a moral attitude even when individuals can provide no rational argument for that attitude, as when they declare that consensual incest between adults is wrong. This does not show that reasoning is absent from moral judgment in all cases, but it suggests that reasoned principles can be absent.

Similar results have been obtained relating moral judgments to anger and other emotions, though these have been less actively explored than disgust because they are harder to induce in laboratory settings. Research in behavioral economics has confirmed that emotions play a role in judgments of fairness and in the motivation to cooperate with others or punish them for defection. Anger and guilt are theorized to play a role in economic exchanges, but the specific identity of the operative emotions is still under investigation.

The finding that emotions influence moral judgments is consistent with views that say moral judgments contain emotions (varieties of sentimentalism) or that judgments are sometimes based on gut feelings (affect-based rules, social intuitionism, and dual-process theories). There is also evidence that a reduction in emotional response leads to a reduction in moral judgments. Simone Schnall has shown that hand washing reduces negative affect and thereby leads to a reduction in judgments of moral severity. Piercarlo Valdesolo and David DeSteno found that watching comedy can make people 3 times more likely to say that it is permissible to kill one man to save five others. The impact of emotion reduction can also be measured in clinical populations. James Blair has shown that psychopaths, who are known to suffer from flattened affect, fail to distinguish between moral and conventional rules. He argues that members of this population see morality as a
set of conventions; without emotions, they cannot recognize that moral rules are very serious and independent of authority.

Blair interprets his results showing that, without emotions, it is impossible to make genuine moral judgments. But defenders of dual-process theories offer another interpretation, suggesting that psychopaths may be incapable of making moral judgments that are emotional in nature but are perfectly capable of making moral judgments based on reasoning. In support of this, Michael Koenigs and Liane Young led a study on people with injuries in the ventromedial prefrontal cortex, a brain area known to play a role in generating negative emotional responses to costly or unpleasant situations. These individuals made moral judgments that looked like those of a healthy control group when they were asked to consider Greene’s track switching dilemma, but they were much more likely than the control group to say it is permissible to push someone in front of a runaway trolley to save five others. Koenigs and Young conclude that emotional impairments may leave some kinds of moral judgments intact while interfering with others. But they admit that this conclusion is only preliminary. People with ventromedial injuries are not completely devoid of emotions, and they also have memories of moral judgments they had made when their brains were intact. We don’t yet know if someone with no emotional responses or no exposure to or recollection of moral judgments made by people with healthy emotional responses could make normal moral judgments.

In summary, research shows that emotions generally occur when people make moral judgments, extraneously introduced emotions increase judged wrongness, and emotional reduction diminishes judged wrongness. This suggests that emotions are integrally involved in moral judgment, as sentimentalism and related theories have claimed. However, there remains an empirical possibility that some moral judgments are made without emotions, as dual-process theories claim, and this possibility leaves room for philosophical theories that say we should limit the role of emotions in moral judgments.

Jesse Prinz

See also Emotion, Structural Approaches; Emotion Regulation; Experimental Philosophy; Moral Development; Rationality of Emotion

Further Readings

EMOTION AND PSYCHOPATHOLOGY

Ever since Emil Kraepelin (1856–1926), the father of modern psychiatry, systematized our understanding of psychopathologies early in the 20th century, clinicians have recognized that diverse cognitive symptoms as well as imbalanced emotional-affective
feelings characterize mental disease, disorders, or problems. Cognitive disruptions are most evident in disorders such as the schizophrenias, which are typically also accompanied by strange feelings and social behaviors. Affective disorders—mood and anxiety problems reflecting brain emotional imbalances—are also accompanied by cognitive changes, including self-blame and obsessive ruminations about one’s hopelessness and worries.

Thus, while schizophrenic disorders are characterized by thinking and social-relational deficits, mood and anxiety disorders reflect disturbances of emotional-affective systems.

There are affects other than emotional ones, such as those reflecting body-state regulations (homeostatic affects) and the pleasures and displeasures of sensations (sensory affects). But these are not as important for understanding psychopathologies.

Although many modern scholars conflate cognitive and affective processes of the BrainMind, they also need to be seen as distinct processes (brain and mind are used interchangeably here, each capitalized, to affirm the prevailing monistic view that neural and mental processes are totally interwoven, and neither concept should have priority). Still, from a brain perspective, important distinctions between cognitions on the one hand and emotions or motivations on the other do need to be made. Cognitions rely critically on brain functions specialized to process information coming into the MindBrain from the outside world. Affective feelings reflect intrinsic, within-brain state-control networks that are shaped, but not created, by external events. Affective feelings are evolutionary gifts for guiding life choices. Unfortunately, these gifts can spoil in untold ways, yielding mental disturbances and experiential distress—the defining characteristic of most psychopathologies. This short essay discusses the emotional foundations of psychopathologies, with a special focus on depression.

For future progress in understanding emotional imbalances underlying psychopathologies, affective BrainMind perspectives are essential. Although the role of mind in behavioral, cognitive, and other psychological sciences remains a contentious issue, psychiatrists and psychotherapists must deal with the shifting tides of affective experiences—the dynamics of the emotional mind—to understand and help people in distress. Skilled psychotherapists know how to use unconditional positive regard, empathy, and frank discussions of emotional lives, yielding insights to help establish new levels of emotional balance in troubled minds. Psychodynamically oriented clinicians often yield the most lasting benefits. Unfortunately, the temperamental characteristics of therapists are typically as important in promoting emotional change as the specific therapies undertaken. Benefits from interacting with caring others, who provide sympathetic attention to the lived details of individual lives, are commonly as effective as psychopharmaceuticals.

Skilled biological psychiatrists often use pharmacological tools to establish affective balance, but a judicious blending of psychological and pharmacological approaches (which often allow lower doses of medications to be used) usually yields optimal results. Ongoing enquiries into the neurochemical nature of brain affective processes, at many levels of MindBrain processing, are bound to yield more powerful and consistently effective drug therapies than currently exist. Indeed, a recent multicenter antidepressant trial (called the STAR*D study) on the efficacy of the most commonly used antidepressants, the selective serotonin reuptake inhibitors (SSRIs), yielded only a 28% overall remission rate: no better than the antidepressant benefits seen from psychotherapy.

Affective Emotional Networks in Psychopathologies

To understand affective experience, normal and abnormal, we need to understand the evolutionary nature of consciousness and the multitiered structure of brain organization. The affective BrainMind has hierarchical levels of control: (a) primary processes, which are evolved intrinsic functions of the brain, (b) secondary processes, which reflect the basic capacity of organisms to learn and thereby link their bodily feelings to environmental events, and (c) tertiary processes, which are higher brain mechanisms that control our capacity to think creatively on one hand and ruminate about our problems on the other.

These systems are nested hierarchies with abundant two-way interactions among the three levels of control. The primary systems generate diverse instinctual emotional behaviors and their corresponding raw feelings. Higher systems regulate lower ones, so all levels are coordinated for effective
joint actions. In one evolutionary view, this may be envisioned, metaphorically, as a hierarchical economic/political system, (a) with many hardworking people (evolutionarily specialized neurons) spread throughout the BrainMind countryside; (b) middle-management people, with additional special skills, assuring commerce flows smoothly and effectively; and (c) on top, the monarchical/pontifical leadership, with abundant powers to dictate how the lower levels operate, especially when times are good. However, when times turn bad (during stress and the resulting emotional distress), the lower systems can revolt, causing chaos in higher leadership chambers. In this way, emotions become disconnected from higher cognitive regulatory functions.

Much evidence indicates that cognitions and affects are normally in seesaw balance so that cognitions can inhibit emotionality while affective intensity can undermine cognitive thinking. Trends toward psychopathological emotional dynamics are commonly evident, in all of us, during the wee hours of the night. People who wake up prematurely, as is common in depression, often do so with “bad thoughts.” They commonly dwell on the negative side of life, partly because higher cognitive regulatory functions are weakened. In part, psychopathology is the failure of higher cognitive functions to regulate these lower, affective ones. When the MindBrain becomes chronically fraught with such tensions, emotional disorders and problems worthy of being labeled as psychopathologies emerge. Ultimately, emotional disorders percolate throughout the nested hierarchies of the BrainMind, with a diversity of nuanced differences that lead psychiatrists to split disorders into subforms albeit with no agreed-upon understanding, yet, of what is broken and imbalanced in the brain.

However, great progress is being made. For instance, mood disorders ultimately arise from imbalances or dysregulations in low-level, primary-process emotion and mood-generating brain circuits. Most current mind medicines help people because they rebalance diverse brain chemistries, especially at the primary-process levels of these multilayered circuit activities. On the other hand, most psychotherapies seek to rebalance emotional lives by providing better regulatory control from above. When psychological and medicinal treatments are combined, they work most beneficially on the affected circuits.

Brain evidence across mammalian species indicates there are at least seven primary-process emotional networks, including reward SEEKING, irritable RAGE, “up-tight” FEAR, loving CARE, distressing PANIC/GRIEF, and euphoric PLAY, as first enunciated by Jaak Panksepp. All these systems are readily studied in animal models. The capitalizations highlight a new way of speaking: They assert the existence of brain networks that govern class-similar emotional behaviors, as well as distinct types of affective experiences in all mammals. This terminology helps minimize semantic confusion arising from the use of vernacular terms, where there is little scientific agreement about the meaning of the words. All investigators who have worked on these systems know these primary-process circuits not only trigger instinctual emotional behaviors but also generate affective feelings, since electrical stimulations of localized areas serve as “rewards” and “punishments” in various learning tasks. However, it is noteworthy that with the advent of human brain imaging, which is more sensitive at detecting changes in higher rather than lower brain functions, many investigators just focus on what in the brain is unique to humans. For instance, neuroscientist Antonio Damasio has emphasized that many cognitive decisions are based on how affects control higher mental processes.

Since arousal of these brain networks yields both instinctual emotional behaviors and corresponding affective states in all mammals, we can use animal models to work out many of the general principles of basic emotions in the human brain. Emotional MindBrain systems provide highly relevant endophenotypes (natural, basic BrainMind functions) for the systematic study and understanding of affects that may be most relevant for understanding disorders such as depression and other affective disturbances. Endophenotypes reflect natural processes of the brain—for example, real networks of the brain—as opposed to conceptual terms, such as most psychiatric categories.

It is especially useful that arousal of these basic emotional systems can be estimated by monitoring emotional vocalizations—distress calls and happy euphoric calls in animal models. This is a large improvement on the traditional use of very general behavioral measures to index depressive changes in animal models—things such as decreased sucrose intake, reduced exploration and social activity, and
diminished struggling in forced swimming tasks. These measures do not have dedicated brain systems that are well understood. In emotional vocalizations, by contrast, we can be sure that there are specific networks that participate in emotional feelings. Hence, we can finally reliably estimate what is transpiring affectively in the minds of animals subjected to various stressors that model many aspects of human depression.

Because of the diversity of emotional networks, there can be many types of emotional problems, all of which, to some extent, reflect imbalances in the endophenotypic primary-process emotional systems of the brain. For instance, the FEAR system is concentrated in midline neural circuits, extending from midbrain hot spots such as the periaqueductal gray (PAG) to primitive forebrain regions such as amygdala, and it generates much of the awful tension of anxiety disorders. Antianxiety drugs such as benzodiazepines effectively quell arousal of FEAR circuitry. Panic attacks, on the other hand, emerge largely from other affective systems, perhaps with precipitous arousals of certain networks that control social connectedness, such as the PANIC/GRIEF system. However, this system, perhaps with chronic arousal, also promotes depression.

Depression, the most common emotional problem in the world, comes in many forms, but all of them tend to be precipitated by chronic stress, especially by the loss of loved ones when one is young and dependent on them. Thus, chronic feelings of PANIC/GRIEF, yielding sustained psychological pain, set the stage for depression in later life. Although transitions from normal emotionality to excessive affect remain to be spelled out empirically, according to theoretical ideas first promoted by John Bowlby, the psychic pain precipitated by prolonged social loss (PANIC/GRIEF overarousal, as currently understood) promotes the emergence of unhappiness and psychological emptiness—the despair and hopelessness which constitute the sustained core of depression. From this vantage point, the key primal emotional networks we must understand better would be intensified PANIC/GRIEF (and to some extent RAGE and FEAR), along with diminished LUST, CARE, and PLAY influences, leading gradually to diminished reward SEEKING, yielding the hopelessness of depression.

Where are these basic emotional circuits situated? Primal emotions and their affects—intrinsic ways to anticipate and deal with life’s challenges—are concentrated in ancient, medial regions of the brain stem, especially in the periventricular regions. Most cognitive functions reflect activities of more recent brain developments, situated farther up (more rostrally) and further out (more laterally) in the brain. The medial location of affect regulation structures remains true in higher regions of the brain as well, regions such as anterior cingulate and frontal cortices, known as the default mode network (DMN), which exhibits elevated activation when people’s attention is not “magnetized” by external events. DMN arousal often presents when people are dwelling on their personal, self-related issues. In depressed people, this frequently reflects rumination on terrible life circumstances and feelings.

Most of the primal emotional systems converge on medial cingulate and frontal emotion-regulating DMN regions, where raw affects blend with cognitive processes. It has been shown that depressed people have unusual activity in these brain regions—overactivity in most individuals but also underactivity in some. Activity in one of these cortical areas (called the subgenual cingulate, or Area 25) is so abnormal that neuropsychiatrists have been able to electrically stimulate this brain region and rapidly reestablish normal feelings in severely depressed patients. These are the highest brain regions to which the PANIC/GRIEF system directly projects, in its course from the midbrain (especially the PAG), through the medial thalamus, to medial forebrain regions.

As negative feelings percolate throughout the higher affective MindBrain, people begin to obsessively dwell on their problems. These ruminations lead the DMN to become superactive and unresponsive to cognitive inputs. This is empirically indicated by external stimuli having less effect on DMN arousal in depressed than in nondepressed people. The despair of chronic depression is also accompanied by dampening of dopamine-mediated SEEKING urges that project to medial frontal regions, thereby promoting feelings of helplessness and hopelessness. That SEEKING circuitry is presently considered to another area that investigators such as Volker Coenen and colleagues feel may be a target for brain stimulation to obtain antidepressant effects in treatment-resistant patients.

Since the PANIC/GRIEF system is intensely inhibited by brain opioids—endorphins and
enkephalins—such feel-good, pleasure-promoting molecules should alleviate depression. They do, and they were once widely used—before modern antidepressants were discovered in the mid-1950s. Since then, safe and nonaddictive opioids such as buprenorphine have been discovered. These opioids become inactive at high doses as they block opiate receptors. At miniscule doses, they effectively alleviate depression, even among those who obtained little relief from other therapies.

In this context it is noteworthy that positive social activities also release brain opioids. Although space does not permit consideration of recent neuroscientific work on the benefits of psychotherapy, it is likely that many benefits arise from the ability of positive human interactions to rebalance the nested hierarchies of various emotional networks of the affective brain.

Jaak Panksepp, Georg Northoff, and Mark Solms

See also Anxiety Disorders; Consciousness, Comparative Perspectives; Emotion, Psychophysiology of

Further Readings


EMOTION AND WORKING MEMORY

In daily life, emotions come and go—often influencing our thoughts, decisions, and actions. Understanding this influence is crucial for developing a comprehensive psychological science. Pivotal to this quest is understanding how emotions interact with working memory, the cognitive system that holds information actively in mind to be worked with and manipulated for reasoning, problem solving, and decision making. Emotion and working memory can interact in three basic ways or modes: (a) emotion may modulate working memory, (b) working memory may modulate emotion, and (c) emotions can be memoranda (i.e., things that are remembered) within working memory. Each mode will be considered in turn.

Emotional Influences on Working Memory

Here we consider emotion as the independent variable and working memory performance—measured by accuracy and speed—as the dependent variable. Intuitively, one would expect a person’s emotional state or mood to influence working memory. Indeed, the science bears this out in intriguing and complex ways. For example, positive emotion has been shown to improve verbal working memory performance, whereas negative emotion impairs it. The opposite occurs for spatial working memory:
Negative emotion can improve performance, whereas positive emotion can impair it. This pattern, known as a double dissociation, aligns well with some models of hemispheric specialization; evidence indicates that language processes and positive affect are lateralized to the left hemisphere, whereas visuospatial processes and negative affect are lateralized to the right. Nevertheless, other studies find minimal effects of emotion on working memory. Inconsistent outcomes could be due to weak levels of emotional arousal and to differences in methodology. Emotional states such as anxiety and emotional disorders such as depression can have a broad-based impact on many cognitive domains, including working memory. Future research will need to investigate the specificity of these effects and how individual differences in affective/personality traits influence working memory.

Working memory performance can also be affected by the emotional valence of the memo-randa. For example, performance may differ when remembering emotional versus neutral words, faces, or pictures. However, the research results depend on the specific process under investigation and on the emotion of the stimulus: Angry expressions may enhance working memory for facial identity, updating the contents of working memory may be slowed by negative words, and interference from prior episodes may be reduced by emotional words. Working memory ability can also be compromised by unexpected emotional stimulation, and these adverse effects may be reduced by the operations of specialized brain mechanisms in the prefrontal cortex.

**Working Memory Influences on Emotion**

The second mode of interaction relates to how cognition can modify, control, or regulate emotion. Here, emotion is the dependent variable. For example, does our susceptibility to intruding emotions depend on how much or how little working memory is preoccupied with some cognitive task? Research on this question is surprisingly sparse. Some evidence indicates that engaging working memory with demanding math problems can reduce negative affect evoked by disturbing pictures. Working memory might be useful for distracting the mind from emotional experiences, subsequently regulating the emotion. Future research is needed to investigate this mode of regulation and whether individual differences in working memory capacity influence emotional responsivity and regulation.

**Affective Working Memory**

The third mode of interaction between emotion and working memory involves emotions themselves as the contents of working memory. Note that distinct from emotional words or faces, emotions as memo-randa are feeling states that are actively maintained even though the eliciting event has passed. This mode motivates the novel concept of a separable working memory subsystem for emotion: affective working memory. This affective subsystem is the counterpart to the well-established verbal and visual subsystems. Although the study of affective working memory is in its infancy, there are data supporting the construct’s viability.

Specifically, evidence favors the operation of separable processes that underlie the ability to actively maintain emotional intensity versus brightness intensity (the latter relying on visual working memory); secondary cognitive tasks can impair visual maintenance and not emotion maintenance, whereas a secondary emotion regulation task can selectively impair emotion maintenance. In a study testing age differences, older adults showed a predictable decline in visual working memory but preserved affective working memory—supporting separable processes. Interestingly, the valence of the maintained emotion affected the age groups differently: Younger adults were better at maintaining negative emotions, whereas older adults were better with positive emotions. Preliminary indications of the brain regions involved with emotion maintenance align well with current knowledge of emotion and the brain: Orbital frontal cortex is recruited when actively maintaining an emotion, and the amygdala, a subcortical structure central to emotion, shows activity that varies depending on the intensity of the emotion being maintained. Future research on affective working memory may have implications for understanding the mechanisms of empathy and its disorders, of reflective thought and rumination, as well as the role of emotions in decision making and the development of wisdom.

**Conclusions**

Emotion and working memory have a complex synergistic relationship that must be investigated.
with multiple approaches. Extant data—although limited—indicate these interactions may be fundamental to understanding the emotion-cognition interface and are an exciting research frontier ripe with many unexplored possibilities.

Joseph A. Mikels and Patricia A. Reuter-Lorenz

See also Emotion, Cerebral Lateralization; Emotion Regulation; Working Memory

Further Readings


**Emotion Regulation**

Emotions serve many important intra- and interpersonal functions. However, emotions also can be unhelpful when they are the wrong type or intensity for a particular context. At such times, emotion regulation is crucial. Emotion regulation refers to any process that influences the onset, offset, magnitude, duration, or quality of one or more aspects of an emotional response. Emotion-regulatory processes may be automatic or controlled, effortless or effortful, and unconscious or conscious. Emotion-regulatory processes may diminish, intensify, curtail, or prolong positive or negative emotion depending on the goal of regulation. Emotions can be regulated extrinsically, as when a parent soothes a crying child, and intrinsically, as when an individual decides to avoid a local establishment for fear of running into a former romantic partner.

Extrinsic emotion regulation has been a major focus in research examining the development of emotion regulation. For example, when a young child encounters a stressful situation and becomes visibly upset, this response is externally regulated when he seeks proximity to a parent or caregiver and subsequently displays lower levels of agitation. Especially early in life, parents have incredible influence over the regulation of their children’s emotions, given the amount of control they have over their children’s daily routines, including academic and social schedules, as well as sleep and eating habits. One important hallmark of development is thought to be the transition from primarily extrinsic emotion regulation to primarily intrinsic emotion regulation (e.g., when a child begins to use self-soothing techniques rather than relying on a caregiver).

Intrinsic emotion regulation falls under the more general umbrella of processes referred to as self-regulation. Contemporary emotion regulation research, which focuses largely on intrinsic emotion regulation, builds on prior work on psychological defenses, as well as stress and coping and mood regulation. However, it may be distinguished from these closely related constructs, as coping specifically targets the stress response and mood regulation targets more diffuse mood states. Successful emotion regulation is thought to be a crucially adaptive skill that protects against the onset and maintenance of mood and anxiety disorders and is considered an important dimension of emotional intelligence. Current research has focused on defining different types of emotion regulation and then comparing and contrasting them in terms of (a) their success in modifying multiple responses associated with the target emotion, (b) the cognitive costs of attempting to use them, and (c) their component cognitive processes.

One prominent theory of emotion regulation is the process model of emotion regulation, which builds on the modal model of emotion. To set the stage for this analysis, the modal model of emotion will be outlined, and the process model of emotion regulation that organizes different types of regulation strategies will be described in detail. Although these models are described in reference to intrinsic emotion regulation, they can be applied to extrinsic emotion regulation as well.
The Modal Model of Emotion Generation

Emotions arise when a situation engages an individual’s attention, has significance in light of his or her current goals, and produces a coordinated, yet malleable, multisystem response. In Figure 1, these properties are presented as the modal model of emotion. This model outlines how an emotion arises over time. The first element is a psychologically relevant situation, which is commonly external. However, relevant “situations” also can be internal mental representations. These situations must be attended to in some way, which allows the individual to assess (or appraise) the situation’s familiarity, valence, and value relevance. If the situation is appraised as relevant and valenced, a loosely coupled set of changes in experiential, behavioral, and physiological response systems occurs. The recursive arrow from one emotional response to the next eliciting situation in Figure 1 emphasizes that these responses often change the situations that prompted them.

Consider a concrete example. Jim is an ambitious young professional who has just stepped off the commuter train when he spots Monique, a colleague who was recently promoted above him. Jim was hopeful that he would be promoted instead, and they have not encountered one another since her promotion was announced. For Jim, this encounter is a situation that has the potential to cause an emotion. Because they are headed to the same office, they greet one another and walk together. As they converse, Jim’s attention is drawn to the fact that he had hoped to get the job, which causes him to evaluate her promotion as a failure to achieve his own goals. At this point, he is experiencing the emotions of sadness, frustration, and perhaps anger. This results in an increase in his autonomic activation and an urge to express some amount of hostility toward Monique. As a result, this might lead to the creation of a new situation: a tense and antagonistic professional interaction, which may eventually trigger a second reaction—for example, shame that Jim wasn’t able to recover from this blow to his ego.

The Process Model of Emotion Regulation

The modal model of emotion (Figure 1) provides a framework for representing the major points in the emotion-generative process at which individuals may intervene to change the emotional response. In Figure 2, the modal model is redrawn, adding five ways that regulation can influence the default trajectory of an emotional response. These five points represent five families of emotion regulation strategies: situation selection, situation modification, attentional deployment, cognitive change, and response modulation. This process model provides a framework for understanding the causes, consequences, and mechanisms underlying various forms of emotion regulation.

To illustrate each family of emotion regulation processes, consider how Jim’s emotional responses might be influenced by each type of regulation as he negotiates through his uncomfortable professional encounter.

Situation Selection

Situation selection requires mental simulation of likely future situations, as well as an understanding of the desirable, appropriate, or unacceptable emotional responses to these situations. Successful use of this strategy involves balancing the projected short-term responses to situations with their long-term consequences, which may have different emotional repercussions.

Jim is employing situation selection if, on hearing that Monique has been promoted, he begins driving to work instead of taking the train. In that case, the potential emotional importance of the interaction is never brought into his attention, never evaluated in relation to his personal goals, and therefore, no physiological or expressive responses occur. The new situation that would have occurred is also altered; if the interaction is avoided, it cannot engender
subsequent shame. However, a different situation is created, and reflecting on his choice to avoid her might have other emotional consequences.

**Situation Modification**

Situation modification refers to adjusting the external environment to meet emotional goals. Situation modification requires the belief that situations are malleable and that one is an effective agent of change in those situations. Although emotional situations can be external or internal, situation modification—as defined here—has to do with acting on the external, physical environment. If situation selection involves overcoming situational inertia, situation modification involves applying force to change the trajectory of the situation’s momentum.

Jim can use situation modification to avoid an emotional response even after he sees Monique on the platform. He could choose to quickly change his path to avoid coming within speaking range of Monique. Just as with situation selection, he avoids directing his attention toward and evaluating the emotional significance of the encounter. Likewise, emotional responses are also avoided, and the subsequent situation is changed, albeit differently than if he had used situation selection.

**Attentional Deployment**

Attentional deployment occurs when individuals direct their attention toward or away from emotional meaning. Attentional deployment can be thought of as an internal situation modification, when the salient aspects of a situation are brought into (or out of) focus. Three commonly used examples of attentional deployment are distraction, concentration, and rumination. Distraction focuses attention on anything besides the potentially emotional aspects of a situation. Concentration is the opposite of distraction and involves focusing attention even more intensely on the emotional features of a situation. Rumination refers to chronic, perseverative focusing, or the continued redirection of attention toward one’s feelings and their consequences, which can prolong an emotional experience.

Imagine that Jim decides to engage Monique in a conversation that leads to the discussion of her promotion. He might then use attention deployment to direct his gaze away from her face, to avoid perceiving the smile that indicates her pride at the promotion. He can also keep his gaze on her face but distract himself by thinking of a pleasant lunch meeting on his schedule, thereby avoiding consideration of the significance of her promotion. This would prevent the evaluation of her behavior as relevant to his personal goals and therefore no emotional responses would result. The subsequent situation will therefore also be changed, given that Jim might not appear authentically engaged in the conversation.

**Cognitive Change**

An emotional appraisal is an assessment of the way the situation relates to an individual’s current goals, and specific appraisals are thought to lead
to distinct emotions. Cognitive change exploits the flexibility of these appraisals, changing their emotional impact by altering the assessment of situations, the relationship of situations to one’s personal goals, or one’s capacity to handle situations. One form of cognitive change that has received particular attention is reappraisal. Reappraisal involves generating and then subscribing to an appraisal other than the one that would have prevailed without intervention. Several studies have shown that the use of reappraisal can successfully modulate self-reported negative affect, startle blink response, and blood flow to the amygdala. If Jim engages in conversation with Monique and attends to the significance of her promotion, he can reappraise the situation by reminding himself that he’ll have other opportunities for a promotion or that his resentment for Monique won’t change the outcome at this point in time. This would reduce the physiological and behavioral aspects of his response and also change the new situation that is created as they converse.

Response Modulation

Response modulation refers to attempting to change one’s physiological, experiential, or behavioral responding directly, once an emotion is already underway. Food, drugs, and alcohol are commonly used to directly manipulate the physiological and experiential aspects of the emotional response. Another typical and relatively well-studied type of response modulation is the regulation of emotion-expressive behavior, including facial expressions, verbal output, and other motor responses. Studies have shown that artificially generating emotion-expressive behavior can increase the experience of that emotion. Directly decreasing emotion-expressive behavior has mixed effects on emotion experience (decreasing positive but not negative experience) and actually increases emotion-related activation of the central and peripheral nervous systems.

Jim might choose to engage with Monique, attend to her promotion, and appraise it as damaging to his career. This would lead to an emotional response, but Jim could still use expressive suppression, hiding his feelings of jealousy behind a mask of indifference or feigned commendation. He could also directly influence his physiological response by smoking a cigarette to relax. Either of these choices would then change the interpersonal situation that is created.

Concluding Comment

Although research on emotion regulation draws from several traditions, the modal model of emotion and process model of emotion regulation characterize emotion regulation strategies by their points of intervention in the emotion-generative process. These distinctions have been quite useful in advancing the understanding of how emotions are successfully regulated by different people in different situations.

Kateri McRae and James J. Gross

See also Attention and Emotion; Emotion, Psychophysiology of; Emotion, Structural Approaches; Emotion and Moral Judgment

Further Readings

The cross-cultural work of Charles Darwin, Paul Ekman, Wallace Friesen, and Carroll Izard has supported the existence of primary universal emotions (i.e., happiness, interest, sadness, fear, anger, contempt, disgust, excitement, and surprise) and cognitively complex emotions (e.g., pride, shame, and embarrassment). Acknowledged in several of these models are affect programs that refer to specific neurological mechanisms mediating the encoding, storage, and output of each emotion type. Based on these affect programs, discrete emotional labels are then derived through an individual’s exposure to a specific culture.

An alternate perspective, however, posits that emotional constructs are better explained through dual-system models of emotion that address dimensions of valence (positive versus negative), arousal (calm versus excited or high versus low), or pleasure (pleasant versus unpleasant). These dimensions also stem from universal or underlying neurological predispositions that are connected to specific learned behaviors (e.g., perception, experience, and expression of specific emotions) via cultural influences. The dimensions can be mapped onto any mode or channel (e.g., facial [face], prosodic [intonation], verbal/lexical [speech content or language], postural [positioning and/or carriage of the body], and gestural [nonverbal communicative movements of the body and/or limbs]) of emotional recognition and described using a wide range of terms depending on the individual culture.

Neuropsychological Theories and Models
Based on early lesion studies, Paul MacLean and James Papez hypothesized some of the first neurological models supporting emotional processes. These models focused on structures within the limbic system as the primary mechanisms of emotional processing but noted the importance of interconnections with the cortex. More recent models of emotion have examined underlying intra- and interhemispheric differences within the brain. For example, Guido Gainotti and Carlo Caltagirone have posited that various neuroanatomical regions mediate certain features of emotional processing based on automatic components versus control of spontaneous emotional expression. In this model, the right hemisphere is considered important for the mediation of the automatic components, and the left hemisphere is linked with the control of spontaneous expression. Dawn Bowers, Russell Bauer, and Kenneth Heilman have posited a somewhat different explanation of these brain-behavior relationships. Bowers and colleagues suggest that right-hemisphere mediation of nonverbal affect lexicons plays a role in the appraisal of emotional stimuli (e.g., face, prosody, and gesture). The lexicons are considered to be a part of a comprehensive cortically based affect-processing network. Last, Joan Borod’s componential model proposes that emotional content is mapped onto various perceptual stimuli via a central affective processor. This central processor functions irrespective of modality and is distinguishable from more finite cognitive functions. Specific cortical and subcortical regions subserve the various components of each emotional process.

Right-Hemisphere Hypothesis
Traditionally, the predominant neuropsychological hypotheses regarding emotion recognition focused on hemispheric specialization (i.e., right hemisphere [RH], left hemisphere [LH], and both hemispheres [bilateral]). Early perspectives supported RH dominance in the recognition of emotion regardless of the target (i.e., emotion, type, or valence). This evidence stemmed from work primarily examining recognition of facial emotion and affective prosody in healthy adults and individuals with unilateral brain damage and psychopathology. Because of inconsistencies in the above findings, alternative models were developed suggesting that emotion recognition is mediated by caudal cortices within the RH. Joan Borod and colleagues reviewed the findings from a number of neuroimaging studies and summarized posterior (i.e., temporal and/or parietal) cortical activation results for emotion recognition in healthy adults. Examining facial and prosodic channels of communication, there were more RH findings than LH or bilateral findings. In contrast, for the lexical channel of communication the distribution of findings was relatively equivalent across these three possibilities (i.e., RH, LH, or bilateral involvement). These neuroimaging findings are consistent with an extensive review of the behavioral findings for emotion recognition in healthy adults by these authors in the same review paper. A somewhat disparate model has been proposed that suggests RH dominance for stimuli engaging preattentive/
unconscious processing of affective stimuli and/or elevated levels of arousal, with LH mediation relegated to the conscious processing of emotion.

**Valence Hypothesis**

By contrast, the valence lateralization hypothesis has posited hemispheric specialization for specific emotion types or dimensions. In the original formulation, the RH was dominant for negative emotions, and the LH was dominant for positive emotions. These findings were based on early studies examining electroencephalography in individuals viewing affective stimuli, behavioral findings for individuals with unilateral brain damage, and mood changes in epilepsy patients undergoing the sodium amobarbital (sodium Amytal) procedure (or Wada test). The Wada test is used to determine which cerebral hemisphere of the brain is critical for particular cognitive and affective functions. Sodium amobarbital is directly injected into an internal carotid artery or introduced via a catheter in the femoral artery and anesthetizes one of the two hemispheres allowing for the assessment of functional lateralization.

A revised version of the valence hypothesis supported RH parietal mediation of perception/recognition of emotion, especially for negative emotion. In this version, anterior brain regions (i.e., frontal lobe regions) differentially mediate experience and expression of emotion (i.e., RH dominance for negative emotions and LH dominance for positive emotions). This latter distinction is more commonly referred to as the dimension of withdrawal/avoidance versus approach.

**Additional Dual-System Models**

Neuropsychological models have also been posited to explain hemispheric differences for experienced emotional processes in relation to motivation. Jeffrey Gray developed the reinforcement sensitivity theory that focuses on two motivational systems: the behavioral inhibition system (BIS) and the behavioral approach system (BAS). This theory is used to describe individual differences in personality traits. Further, it explores how related emotional experiences are based on the interaction between neurobiological predispositions and environmental cues. The RH has been associated with the mediation of the BIS and the LH with the BAS. Another dual-system model proposed by Elliott Ross and Ross Buck ascribes hemispheric differences based on a conceptual distinction between a social versus a self-focused affective system, mediated by the LH versus RH, respectively.

**Methodology**

**Traditional Experimental Paradigms**

During the past four decades, a number of experimental paradigms have been utilized to examine the recognition of emotion in healthy individuals and clinical populations (e.g., mood disorders, schizophrenia, stroke, and seizure disorders). Traditional neuropsychological methods have included a variety of behavioral laterality techniques (e.g., dichotic listening, tachistoscopic viewing, and free-field facial viewing).

In the typical dichotic listening task, participants are presented simultaneously with two different auditory stimuli (e.g., emotionally intoned words, nonsense syllables, or nonspeech sounds), one to each ear, and are asked to either passively listen to the stimuli or attend to specific components of the stimuli. Participants are then asked to respond in some fashion to what they hear (e.g., repeating sounds/words or answering questions regarding the content of the stimuli). In the tachistoscopic viewing task, participants are presented with visual stimuli (e.g., facial expressions or words used to convey emotion) to the right and/or left visual fields for a very brief period of time. The responses provide information regarding lateralization for emotional functions. In the free-field viewing procedure, developed by Jerre Levy, individuals are asked to view chimeric faces in both the right and left hemispaces and make a judgment regarding the face (e.g., congruency of the facial expressions or the hemiface expressing more emotion). There are several approaches to creating chimeric faces; in some emotional paradigms, stimuli are made by combining photographs of an individual producing two different types of facial expressions. The bisected expressions are joined at the vertical midline.

Many of the studies using facial stimuli have focused on static/still photographs and have asked participants to match, discriminate, or identify specific emotions. Dynamic emotional displays have been used to capture emotion recognition as it might
occur in real-life situations. Similarly, investigations examining emotion recognition through prosodic and verbal modalities have used static and dynamic stimuli in discrimination and identification paradigms. Opinion differs as to the optimal method for assessment, in part, because disparate findings have occurred using similar paradigms and stimuli.

**Emotion Test Batteries**

Clinically, a number of neuropsychological test batteries are available to assess emotional processes using behavioral techniques. Almost all these batteries include several of the primary emotions (e.g., happiness, sadness, and/or anger) and a neutral condition, as well as tasks evaluating the recognition of emotion across single or crossed modalities (i.e., meaning stimuli presented in two different modalities, for example, visual and auditory). Tests evaluating emotional expression and experience, in addition to tests of recognition, are included in several of the batteries. However, psychometric support for these measures, especially related to construct validity and ecological validity, is limited at the present time.

The tests of emotional recognition for each battery are described below, and the batteries are presented in alphabetical order. The Aprosodia Battery evaluates comprehension of affective prosody employing identification and discrimination of stimuli using asyllabic and monosyllabic utterances, as well as words. The Battery of Emotional Expression and Comprehension includes tests measuring perception of emotional faces, prosody, and scenes. The Comprehensive Affect Testing System allows assessment of discrimination, identification, and cross-modality matching of facial and prosodic emotions. The Florida Affect Battery includes subtests evaluating the perception and recognition of nonverbal emotional signals via facial and prosodic channels, as well as a nonemotional discrimination subtest. The New York Emotion Battery examines perception (identification and discrimination) across facial, prosodic, and lexical channels using eight emotions divided across positive/negative valence and approach/withdrawal dimensions. Nonemotional control tasks are included as a feature of this particular emotion battery. The Perception of Emotions Test assesses emotion across facial, prosodic, and verbal modalities using a multiple-choice recognition format (i.e., four drawings) and the corresponding emotional label. The Victoria Emotion Perception Test evaluates the ability to identify and discriminate emotional categories and intensity levels through facial and prosodic channels.

**Neuroimaging Techniques**

The rapid emergence of neuroimaging techniques has facilitated more direct evaluation of brain-behavior relationships and the acknowledgement that a variety of specific neuroanatomical structures/subsystems and neurophysiological processes are activated during most emotion recognition paradigms. In these studies, participants are typically exposed to discrete emotional stimuli within a specific modality/channel (e.g., faces, voices, or words) while neural and, in some cases behavioral, responses are recorded. Participants may engage in explicit recognition (make an explicit judgment of an emotional stimulus) or implicit recognition (passively view emotional stimuli). The task demands and approaches to analyzing the data are quite variable, resulting in some degree of difficulty comparing the findings from the various studies.

Neuroimaging techniques have allowed for examination of the temporal course (i.e., electroencephalography [EEG], event-related brain potentials [ERPs], and magnetoencephalography [MEG]), and the spatial resolution (i.e., single-neuron recording, functional magnetic resonance imaging [fMRI], positron emission tomography [PET], single photon emission computerized tomography [SPECT], and transcranial magnetic stimulation [TMS]) of emotion recognition within the brain.

Neurophysiological techniques, including EEG and ERP measures, evaluate the summed electrical activity of neuronal groups on the scalp during rest and in response to tasks, respectively. fMRI, PET, and SPECT isolate regional activation by examining changes in levels of hemodynamic or metabolic functions. These latter techniques examine activity/changes in blood flow and metabolism of ions in response to specific task performance. TMS allows for the application of pulsed magnetic fields over specific areas of the head and into the underlying neural anatomy. Similar to direct electrical stimulation of brain tissue, TMS can result in temporary inhibition and excitation of brain areas and in subsequent emotional processes when the device is activated.
Functional Neuroimaging Studies

The current neuroimaging literature as applied to emotion recognition has focused on the facial channel with a paucity of work existing for prosodic and lexical channels. The majority of the following discussion is based on work using fMRI techniques.

Facial Channel

Neuroimaging paradigms typically either assess differential neural activation in response to recognition of different facial emotional expressions or in response to neutral versus emotional facial expressions. The findings from several neuroimaging studies using facial stimuli support the idea that there are specific neuroanatomical substrates in the RH that are involved in emotional recognition. Ralph Adolphs, Hanna Damasio, Daniel Tranel, and Antonio Damasio asked participants with focal brain lesions to judge facial expressions of six basic emotions. Using a novel technique, anatomical descriptions of surface lesions and task performance scores were jointly mapped onto a standard brain space. The regions that best correlated with impaired recognition of emotion were within the right inferior parietal cortex and the right mesial anterior infralinear cortex. No impairments in facial emotional recognition were noted in the LH damaged patients. Additionally, several discrete neural mechanisms, which are not necessarily lateralized, have been identified in response to facial stimuli, especially within the temporal and frontal areas. Regions in visual cortices, face selective areas of the fusiform gyrus, the amygdala, and the orbitofrontal cortex exhibit greater neural activation to passive viewing of emotional than neutral faces. Several studies have found that the enhancement of activations within the fusiform cortex and amygdala in response to emotional facial expressions persists with reduced attention or awareness, as for example, when faces appear at task-irrelevant or ignored locations. The findings from a study conducted by Patrik Vuilleumier and Gilles Pourtois suggest that the visual system may prioritize emotional stimuli over neutral stimuli; in fact, emotional faces may capture attention in a reflexive manner. Furthermore, enhanced responses to emotional versus neutral faces have been observed in the following regions during explicit processing: the orbitofrontal cortex, ventromedial prefrontal cortex, anterior cingulate cortex, and posterior superior temporal sulcus, as well as somatosensory areas and insular cortex.

Substrates for Basic Emotions

Recent research is also being conducted to identify dissociable neural substrates for recognizing individual basic emotions. Neuroimaging results provide evidence of a dissociation for fear and disgust. Whereas the insular cortex and basal ganglia are involved in the recognition of disgust but not fear, the amygdala is involved in the recognition of fear, but not disgust. Regarding the recognition of anger, the cingulate, orbitofrontal cortices, and the amygdala have been isolated via neuroimaging studies. Some evidence further suggests that certain neural regions may be particularly sensitive to characteristic facial features of a basic emotion. For example, some findings support the notion that viewing the eye region of a face producing fear is sufficient to produce activation in the amygdala. Regarding positive emotions (i.e., happiness), the orbitofrontal cortex, anterior cingulate cortex, basal ganglia, and amygdala have been connected to the recognition of happy faces. Furthermore, it has been shown that activity in the left amygdala decreases with increasing happiness. Additional research is needed to validate claims of specific neural substrates for the recognition of happiness and other emotions, including anger, surprise, and sadness.

The Amygdala and Recognition of Fearful Facial Expressions

The importance of the amygdala and associated neural regions for the identification of facial expressions of fear has been a consistent finding in the literature. Conscious and unconscious perceptions of fearful faces consistently elicit significant amygdala responses. Amygdalar activation has been shown to increase when viewing increasingly intense fearful facial expressions. Interestingly, enhanced activation of fusiform regions in response to fearful faces, compared to neutral faces, is reduced in patients with amygdala lesions. Activation is observed even when using a masked paradigm to present fearful faces, that is, when fearful expressions are not consciously recognized. In a masked paradigm, a fearful face is presented for a very brief interval and then overlaid
with a neutral face so that the participant does not have time to consciously recognize the fearful face.

As a result, much research is currently aimed at understanding what modulates amygdalar activation. Many researchers theorize that more cognitively demanding tasks (such as those requiring labeling emotional content) reduce activation in the amygdala. One study conducted by Ahmad Hariri and colleagues, for example, observed a reduction in amygdala activity (relative to simply attending to stimuli) when participants reappraised negative pictures with various cognitive strategies. Similarly, Stephan F. Taylor and colleagues found that activation in the extended amygdala elicited during different emotional tasks could be modulated by task demands and/or cognitive factors. (By extended amygdala, the authors are referring to the bed nucleus of the stria terminalis and its sublenticular extension into the centromedial amygdala.)

Although many of these effects are robust and consistent across studies, they must be interpreted with caution as the amygdala has also been linked to the recognition of faces portraying sadness and happiness. Additionally, the amygdala is implicated in the response to non-facial displays of unpleasant emotions, including unpleasant auditory, olfactory, and gustatory stimuli. Thus, it is plausible that the amygdala, as a region sensitive to emotional expressions, plays a general role in processing salient stimuli. Consequently, in examining associations between brain regions and specific emotions, it is important to question whether individual emotions are subserved by dissociable neural subsystems.

**Prosodic Channel**

Behaviorally, certain emotions are more difficult to identify via the prosodic channel compared with the facial channel (e.g., disgust). Emotional prosody tasks can be categorized as either implicit or explicit recognition tasks. Typically, in such experiments, participants passively listen to affective speech or engage in an active task (e.g., identify the emotion expressed by the tone of voice or make a judgment about the tone of voice). Similar to emotional facial recognition, it is hypothesized that simple cues, such as frequency range or amplitude variation, can be used to identify the emotion represented in a stimulus.

A number of imaging studies have shown that the RH is important in emotional prosody recognition. Right lateralized processing of the voice occurs at the level of the auditory cortex. Additionally, right frontoparietal regions have consistently been identified as structures involved in emotional prosody recognition. However, although the RH seems disproportionately involved in this process, there is substantial evidence in the literature to support bilateral processing for emotional prosody recognition, especially in orbital regions of the prefrontal cortex and in the amygdala.

Less is known about the specific neural correlates of emotional prosody recognition. However, Dirk Wildgruber has posited a prosody network—that is, a processing model in which there are parallel projections from the posterior superior temporal sulcus to frontal cortical regions. He has examined the neural substrates underlying successive processing stages of meaningful emotional prosody. In one study, participants identified and named the emotion expressed by vocal tone. Findings supported activation within the right inferior frontal cortex and the right posterior superior temporal sulcus. A separate discrimination task compared linguistic prosody versus emotional prosody. Explicit judgment of linguistic prosody (e.g., question intonation) was associated with greater activation in the left inferior frontal gyrus, whereas evaluation of emotional prosody was associated with activation of bilateral orbitofrontal cortex.

Data exist to support the notion that the implicit processing of unattended emotional prosody affects the explicit judgment of facial expressions. For example, it has been shown that the middle section of the right fusiform gyrus shows significantly stronger activation when facial expressions are displayed in the presence of a fearful voice compared to a happy voice.

**Lexical Channel**

There is a paucity of neuropsychological studies investigating emotion recognition in the lexical channel. Paradigms typically incorporate emotional decision tasks by requiring the selection of particular pleasant, unpleasant, or neutral words. Matthias Tabert and colleagues instructed participants to select the most unpleasant or neutral word from a lexical set. Activation in the right amygdala was greater for unpleasant than neutral word sets, and the amygdala showed a sustained response to the
evaluation of unpleasant words. Additionally, the overall peak blood oxygenation level-dependent (BOLD) response in the occipital cortex was strongly positively correlated with the peak BOLD response in the right amygdala in response to unpleasant words but not neutral words.

**Developmental Processes**

The ability to recognize emotion begins in infancy and continues to develop across the life span. Infants as young as 4 months are able to discriminate among simple facial emotions. The ability to perceive and label emotions, which may reflect sophisticated abilities to recognize subtle nuances, blended emotions, and/or less intense emotions, appears to increase during childhood and adolescence. The development of these emotional processes is suberved, in part, by maturation of specific brain structures (e.g., amygdala and prefrontal cortex) and hemispheric subsystems (e.g., frontal-limbic connections) throughout adolescence and early adulthood. This neuroplasticity most likely occurs via neural mechanisms, including myelination, synaptogenesis, and dendritic arborization.

Explicit recognition of facial affect, especially for negative emotions (i.e., anger and sadness), appears to decline with increasing age. Conversely, the recognition of disgust and positive emotion improves with age. Although atrophy in structural volume (e.g., frontal and temporal cortices, basal ganglia, and amygdala) may explain these changes for negative emotions, it does not provide support for the effects on positive emotions. This latter finding may result from the increased social and cognitive adaptations and experiences that accompany the aging process, which strengthen existing subcortical and medial prefrontal subsystems.

**Gender**

Behaviorally, women have been reported to recognize emotional stimuli better than men as well as be more emotionally expressive. One possible explanation for this finding has been that women respond to such tasks by attending to the subjective components of the emotional experience, whereas men respond to the same tasks by directing attention to the sensory and activation components. Several neuroimaging studies of emotional processes have provided limited support for gender-specific lateralization of emotion. These studies have identified increased activation in specific brain regions, and gender effects are more frequently associated with the valence dimension. Tor Wager and colleagues applied a meta-analytic approach to examine the relationships between brain activity, valence, and gender. Their findings did not support global RH lateralization for emotion and provided minimal support for lateralized frontal cortical mediation of the valence dimension. Further, their analyses revealed that women more frequently demonstrated increased activation in the medial frontal cortex, thalamus, and cerebellum during affective paradigms. Activation in inferior frontal cortex, dorsal striatum, and posterior sensory and association cortex was more common for men.

*Lawrence H. Pick, Erica P. Meltzer, and Joan C. Borod*

**Author’s Note:** This work was supported, in part, by Professional Staff Congress—City University of New York Research Awards 62527-0040 & 63746-0041 and by NIH R01 DC01150 subcontract to Queens College.

**See also** Emotion, Cerebral Lateralization; Emotion, Structural Approaches; Facial Expressions, Emotional

**Further Readings**


It can seem obvious that there are natural connections between emotional states of mind, such as being angry or fearful, and their conscious properties, such as feeling anger and fear. However, there is currently no agreed-on or settled understanding of the precise relationships between these phenomena. There are a number of reasons for this. A major factor is the plethora of competing accounts about the nature of emotions, which promote different views about the importance of consciousness, and specific types of consciousness, for understanding the emotions. To give a sense of these options, this entry will describe features of the purer forms of intellectual and experiential accounts.

Emotions have been variously identified with judgments or feelings of bodily changes, and they are sometimes thought of as more structurally complex mental states that, at least ordinarily, involve the former as elementary constituents. Whether consciousness is necessary or matters critically for understanding emotion, and which variety or form matters, depends on the account of the emotions adopted.

### Intellectualist Accounts

#### Propositional Attitudes

There is a long-standing and still popular intellectualist trend to treat emotions as nothing but a species of propositional attitude. Propositional attitudes are psychological attitudes of various kinds that relate thinkers to propositions. The content of such propositions is paradigmatically expressed in natural language sentences by *that*-clauses or content-clauses such as, “The train leaves Euston at 9:15 p.m.” For example a thinker, X, can believe, desire, hope, fear, or recognize p, where the mental state verb denotes X’s attitude and p denotes the propositional content to which X’s attitude relates. Propositions may be true or false—for example, things may or may not be as the thinker takes them to be. As such, propositional attitudes are thought to be states of mind that possess or relate to truth-evaluable contents, where the term *content* can refer either to what is thought about—for example, a state of affairs or object, whether real or imaginary—or the specific manner or way in which that subject matter is thought about. Depending on one’s view of the nature of propositions, propositional attitudes will be regarded as relating a thinker directly to some actual or possible state of affairs or indirectly via a specific mode of presentation or representation. But either way, if emotions are thought of as a species of propositional attitude, then it is an easy matter to explain how such states of mind enter into rational commerce with cognitive states of mind, such as beliefs and desires.

Understood in this way, emotions are akin to judgments, thoughts, or contentful evaluations. Importantly, their authors need not be explicitly aware of making such evaluations. For example, John may be furious at Jane for slighting him even if he has misrepresented the facts and even if he is unaware that he is in such a state of mind. If emotions are nothing but propositional attitudes then, like others of their kind—such as beliefs and desires—they might be tacitly possessed. In this case,
there would be no essential or necessary connection between emotions and consciousness, at least with respect to this understanding of being consciousness.

**Unconscious Attitudes**

This sort of cognitivism about emotions can seem to fit well with psychoanalytic accounts, which invoke talk of unconscious and repressed emotions. For although the psychoanalytic tradition regards emotions as complex states of mind, with both contentful and affective properties, it also emphasizes that the contentful aspects of emotional states are precisely those that become inaccessible to consciousness in cases of repression. Thus, while emotions still play causally efficacious roles in influencing behavior, their meaning or focus may not be available or fully present to their owners. In such cases, it seems agents systematically lack what is known as *access consciousness* to the content of their repressed emotions. Such states of mind would be regarded as unconscious precisely because the agents in question are unable to manipulate their emotions by relating them to other mental states so as to rationally influence or modify their behavior.

**Purely Experiential and Hybrid Accounts**

There is however another form of consciousness that appears to be more intimately connected with the emotions. Consciousness is sometimes used to denote a state of mind with a characteristic feel—that is, a state of mind in which there is something-that-it-is-like to occupy it. Mental states that possess this property are said to be phenomenally conscious; they have a certain phenomenal or qualitative character. The phenomenal character of such mental states depend on a range of factors, such as the kind of being that enjoys them, the kind of stimuli and circumstances that elicit them, and so on. As such, experiencing a feeling of itchiness is quite different from experiencing elation or sorrow. While it is debatable whether or not most propositional attitudes, such as beliefs or desires, possess phenomenal qualities, it is extremely plausible that emotions typically do and arguable that they must do so, essentially.

Emotions might be thought to necessarily or integrally require phenomenal consciousness because to be an emotion of a certain kind necessarily requires the having of certain kinds of feelings. By these lights, nothing would count as an emotion proper if it lacked phenomenal qualities altogether. One way to make sense of this proposal is to imagine that phenomenal qualities are part of the constitutive content of a propositional attitude. Different theorists attempt to accommodate this idea in different ways. Thus the phenomenal properties could be thought of as features (of the external world or oneself) at which subjects are intentionality directed, or they might be thought to be exhausted by representational properties that explain the intentional directedness at such features. Taking these ideas to their extreme, some have argued that emotions are not best understood as being a species of sophisticated propositional attitudes but rather as being essentially perceptual states of mind. If so, the core of emotional responding might be understood in a quite different way than the intellectualist tradition assumes—perhaps, depending on the understanding of perception one adopts, without necessarily requiring emotions to be thought of as content-involving mental states at all.

*Daniel D. Hutto*

**See also** Access Consciousness; Intentionality of Emotion; Unconscious Emotion, Psychological Perspectives

**Further Readings**


**Endowment Effect**

People tend to demand more to give up an object than they would be willing to pay to acquire the same object. That is, people value objects that they happen to own—objects that are part of their endowment—more than identical objects that they happen not to own. This is known as the endowment effect.
This entry will first review empirical evidence for the endowment effect and then discuss its practical and theoretical implications as well as the proposed psychological explanations for the effect.

Evidence and Implications

The endowment effect has been widely replicated with many different objects in many different cultures and has even been demonstrated in nonhuman primates. In one study, students who were randomly given a coffee mug emblazoned with their university logo would not sell their mug for less than $5.25, whereas students randomly assigned to the role of buyer would not pay more than $2.50 for the mug, a price discrepancy that emerged even though both buyers and sellers had previously learned how to make efficient transactions for tokens in an experimental market. In another study, some people were given coffee mugs and asked whether they would sell the mug for various amounts of money; others were simply asked to choose between receiving the mug and receiving various amounts of money. The sellers assigned a median value of $7.12 to the mugs, whereas choosers assigned a median value of only $3.12.

The endowment effect has important consequences for everyday consumer behavior. For example, the endowment effect produces undertrading in markets even with zero transaction costs. To illustrate the concept of undertrading, consider an experimental market in which a random half of the people are given chocolate bars and asked if they would trade their chocolate for a coffee mug; the other half are given coffee mugs and asked if they would trade their mug for chocolate. According to standard economic assumptions, 50% of the people should trade, given that, on average, half of the people would have to trade to match the overall distribution of preference for mugs and chocolate. However, the endowment effect reduces trading rates to approximately 25%, presumably because the objects are more valuable to people who happen to own those objects than to those who do not.

The endowment effect also has important theoretical implications for economics and behavioral decision theory. In particular, the endowment effect violates assumptions of standard economic theory that people act on well-defined preferences that are consistent across measurement context. Such assumptions imply that people who happen to be endowed with an object should demand no more to sell their object than buyers or choosers would be willing to pay to acquire the same object and that an object’s value (i.e., how much money an object is worth) should be independent of the way value is measured (i.e., whether one is asked to state a selling price, buying price, or choose between an object and money). In contrast with such assumptions, the endowment effect implies that people’s preferences are constructed in the context of measurement and are contextually contingent.

Theoretical Explanations

Like many robust psychological phenomena, the endowment effect is multiply determined. One common explanation is that when trading, owners view objects in their endowment as losses whereas nonowners view those objects as gains. Because people are more sensitive to losses than to gains—for example, losing $20 feels more intensely negative than gaining $20 feels intensely positive—owners assign more value to objects and hence demand greater compensation for their loss, compared with buyers. The psychological basis for this loss aversion may be that owners tend to focus on different attributes than buyers when valuing objects in their endowment. Specifically, both owners and buyers tend to focus on and place relatively greater weight on what will be foregone in a given exchange. For owners, the endowed object will be foregone and is therefore attended to and valued more; for buyers, in contrast, the expenditures will be foregone and are therefore attended to and valued more. This attention based explanation helps explain why the endowment effect is reduced by encouraging people to attend to the non-focal attribute; that is, owners focus on money to be gained and buyers focus on the object to be gained.

Another explanation is that owners value objects that happen to be part of their endowment because they see those objects as extensions of themselves and they are motivated to view themselves favorably. This explanation, grounded in self-enhancement tendencies, is most often associated with the “mere ownership effect,” a phenomenon highly similar to the endowment effect, emphasizing personal liking of an owned object rather than valuing or pricing of an owned object.
Whatever the processes underlying the endowment effect, people have limited awareness of how the endowment effect influences their own or other people’s behavior. When nonowners predict how they would behave if they were owners, they underestimate how much they would value the objects they own and how much actual owners value the objects they own. Similarly, owners overestimate how much they would value their objects if they did not own them and how much actual buyers value those objects. The endowment effect is thus pervasive and potent but nonobvious.

*Jacob Westfall and Leaf Van Boven*

See also Allais Paradox; Decision Theory, Philosophical Perspectives; Dictator Game; Dutch Book Arguments; Neuroeconomics

Further Readings


**Envy**

This entry concerns the social emotion of envy, the often-painful mix of displeasure and ill feeling triggered by the awareness of another person’s superiority or advantage. After addressing two definitional distinctions, the entry will address the universal and adaptive nature of the emotion, the range of situations in which people feel it, the hostile nature of malicious envy, the transmutational nature of the emotion, and the implications of envy for happiness.

**Benign Versus Malicious Envy**

Scholars are careful to point out two important definitional distinctions that are often a source of confusion about the nature of envy. First, the word *envy* has two meanings, one referring to a benign feeling akin to admiration, and the other referring to a hostile, more aversive state. As research by Niels van de Ven and others shows, people admit to benign envy, and the feeling seems relatively unconnected to ill will or aggressive behaviors. Hostile envy, scholars assume, is usually denied in public, and even in private, and is closely linked with a variety of aggressive inclinations. Benign, admiring envy is an important emotion to understand, but hostile envy is the prototypical variety and the kind that has been the focus of literary, philosophical, and religious scrutiny since antiquity.

**Envy Versus Jealousy**

The second definitional distinction contrasts envy with jealousy. Generally, envy is dyadic. It involves two people and arises when one person lacks an advantage enjoyed by the other. Jealousy is triadic, and it arises when one person fears losing or has lost the attention of a second person to a third person. As studies by Gerrod Parrot and Richard Smith show, these situational differences associated with each emotion typically lead to distinctive affective experiences. Unfortunately, the word *jealousy* can be used to denote either the emotions of envy or jealousy, which encourages the sense that they are equivalent experiences.

**Universal and Adaptive Nature of Envy**

It is easy to understand why envy is such a prevalent, pan-cultural emotion. People who are superior on valued attributes reap greater power and attention as well as higher self-esteem. Inferiority leads to less power and attention as well as lower self-esteem. It would be bizarre for other people’s consequential advantages to have no emotional effect on us. Furthermore, from an evolutionary perspective, as evolutionary psychologists Sarah Hill and David Buss argue, it would hardly be adaptive to be inclined to feel that another person’s advantage is a fully satisfactory outcome. In this sense, a capacity to feel envy, although it has an unpleasant edge to it, serves a necessary adaptive function. If we,
as evolving beings, had failed to develop an emotion designed to help us keep up with the Joneses, perhaps we would have withered away on the evolutionary vine.

**When Do We Envy?**

Envy is not an invariable reaction to noticing another person’s advantage. Empirical work confirms the insights of Aristotle by showing that people envy those who are similar to themselves. Writers envy other writers rather than acclaimed athletes, for example. This similarity causes the offending comparison to hit home and enables people to imagine what having the advantage themselves might be like. Yet it is a frustrated sense of what is possible. The imagined taste is there but with no real sense that the desired advantage will actually come one’s way. Now-classic empirical work by Peter Salovey and Judith Rodin also shows that people envy others who have advantages on domains that are linked to their self-worth.

**Envy and Aggression**

It is important to highlight that scholarly traditions link envy of the hostile kind with aggression. One can find this link in religious thinking (Cain’s slaying of Abel), analyses of historical events (the Nazis’ hatred of and desire to annihilate the Jews) and classic fiction and drama (Shakespeare’s depiction of Caesar’s assassination, led by the envious Cassius). Empirical work supports this link as well. Studies in experimental economics by Daniel Zizzo suggest that envy may cause people to give up their own highest outcomes as long as an envied other’s outcomes will suffer relatively, thereby “burning the rich.” Also, some empirical work demonstrates that envy can cause people to feel *schadenfreude*, or pleasure at another person’s misfortune, if an envied person suffers. This evidence points to the many undesirable consequences of hostile envy. However, the hostile aspect of envy may serve an especially valuable function by creating an effective incentive for self-seeking behavior. In the long run, failing to seek out prospects for superiority and the advantages that result from superiority may have led to reproductive oblivion, as noted above.

**Repugnant and Transmutational Nature of Envy**

Envy, especially in its malicious form, is socially repugnant. People who admit to their envy are also admitting that they are feeling hostile and that they are inferior in some way. Envy is also threatening to the envying person’s self-image. For these reasons, people feeling envy often hide their envy and are motivated to rationalize their hostility. A typical strategy is to construe the envied person’s advantage as unfair. In fact, a sense of resentment seems a common feature of envy, though this type of resentment should be distinguished from feelings of injustice based on obviously unfair advantage. In general, scholars claim that envy is likely to be repressed, suppressed, or transmuted such that the feeling is given a more socially appropriate label, sometimes fooling observers and the self as well. Envy-inspired hostility may tend to be indirect, taking the form of behaviors such as gossip and backbiting or of feelings such as *schadenfreude* when the envied person fails.

**Envy and Unhappiness**

It has long been claimed that envy is a frequent cause of unhappiness. Comparative standards for self-evaluation are prescriptions for discontent because there will always be someone who compares better. Scholars suggest that an envious disposition works against seeing the intrinsic value in praiseworthy things and marvelous attributes possessed by others. Accomplishments that deserve to be celebrated may be, instead, cause for a sour, invidious mood. Also, help from others may be resented because it suggests one’s inferiority—and the helper’s superiority. Indeed, research by Robert Emmons and Michael McCullough show that feelings of gratitude, so often associated with positive and healthy interactions with others, are largely missing from the envious person’s emotional palate. Envy can be understood as a kind of poison in the system that precludes positive emotions and fosters rancor, bitterness, and resentment. Envy may be an inevitable, adaptive human emotion, but it is unlikely to be an emotion to cultivate. Otherwise, the chances are that happiness goes out the window.

*Richard H. Smith*

*See also* Emotion Regulation; Jealousy; Resentment

**Further Readings**

Event Memory, Development

This entry covers changes in both neural processing and behavior that are related to developments in event memory from infancy through the school years. The study of event memory from a developmental perspective entails investigation of age-related changes in memory for ordered sequences that consist of actors, actions, and objects interacting with each other to achieve a goal. Event memory emerges by the end of the first year of life and undergoes substantial development for years thereafter. Age-related changes are due to a number of factors, ranging from developments in the neural substrate of memory to the social-cultural environment in which development takes place. The result is both normative age-related changes and individual and group differences.

Until the 1970s, it was thought that children under 2 years of age lacked the ability to remember the past. Even beyond infancy, children were thought to have poor memory abilities based on their lackluster performance on laboratory tasks, such as word and picture list learning. New methods and approaches developed since that time have revealed substantial memory abilities even in infancy. Similarly, when they are asked to recall meaningful material, preschoolers provide relatively detailed and organized reports. Application of neural imaging methods (e.g., event-related potentials or ERPs and functional magnetic resonance imaging or fMRI) complement behavioral paradigms (e.g., imitation, verbal reports) and illustrate a protracted course of memory development from infancy through childhood and into adolescence.

The Neural Substrate of Event Memory

The memory processes of encoding, consolidation, storage, and retrieval rely on a network of brain areas. The structures themselves, as well as the connections between them, have different rates of development. The temporal lobe structures implicated in the encoding and consolidation of memory traces (e.g., hippocampus and surrounding cortices) develop from infancy through the preschool years, with less pronounced changes thereafter. The prefrontal cortex, which is involved in retrieval, undergoes a longer period of development, through adolescence. The time course of development of this temporal cortical network is consistent with known behavioral changes in event memory.

Event Memory in Infancy

In infancy, imitation-based tasks are used to assess event memory. Props are used to produce novel sequences of action that result in a goal or outcome. Infants are encouraged to imitate the sequences immediately, after a delay, or both. Use of this task has revealed age-related changes in the robustness, reliability, and temporal extent of memory throughout infancy. Infants as young as 6 months can reproduce the actions of sequences after 24 hours. By 20 months, the robustness of recall increases such that infants recall the actions of events after as many as 12 months. There also are increases in recall of the temporal order of sequences. After a delay of 1 month between encoding and retrieval, 50% of 9-month-olds, 75% of 13-month-olds, and 100% of 20-month-olds recall sequences in correct temporal order. Over this same space of time, ERPs reveal age-related differences in both encoding and consolidation of memory traces, with older infants exhibiting more robust encoding and more successful consolidation relative to younger infants. Differences in ERPs correlate with variability in behavior.

Event Memory in the Preschool and School Years

Age-related changes in event memory continue throughout the preschool and into the school years. Based on their verbal reports, it is clear that children as young as 3 years of age form, retain, and later retrieve memories of past events. In fact, younger children provide the same amount of accurate information about their experiences as older children, even after long delays. Their recall is more dependent on scaffolding in the forms of questions, cues, and prompts, however. Moreover, across the preschool...
and into the middle childhood years, event memory reports become more elaborate, complex, and complete. They also become increasingly evaluative and self-reflective and thus more autobiographical. These age-related changes may be associated with development of the temporal-cortical memory network that continues throughout childhood and into adolescence. Neuroimaging techniques, including fMRI and ERPs, are being used to examine relations between developments in the neural substrate of memory and behavioral changes in recall.

**Individual and Group Differences in Event Memory**

Substantial variability has been observed in the quality and quantity of memory reports, especially for events that are personally significant or autobiographical. Gender differences are observed as early as the elementary school years. Girls tend to produce longer, more coherent, and more detailed narratives than boys of the same age. In addition, girls include a greater number and variety of emotion words in their memory reports. Socialization practices, especially maternal reminiscing style, are a strong predictor of gender and individual differences in children’s memories. Reminiscing style also differs across cultures. For example, Asian mothers are less elaborative than American mothers. In turn, the autobiographical narratives of children from Asian cultures are less detailed. These gender and cultural differences are reflected in the timing of the offset of infantile or childhood amnesia (the relative paucity among adults of memories from the first years of their life and the sparse distribution of memories until 7 or 8 years of age): Women and individuals from Western cultures tend to have earlier first memories. Cognitive and neural factors also are implicated in the explanation of childhood amnesia. For example, use of the cue word technique (reporting a memory in response to a word such as *ice-cream*) to examine the distribution of autobiographical memories suggests different rates of forgetting of events by children and adults. Differential rates of forgetting, in turn, may be related to developmental differences in the temporal-cortical network that supports event memory. Despite childhood amnesia, some events are remembered over long periods of time. Emotionally charged events, such as painful medical procedures and natural disasters (e.g., tornadoes), are remembered years later, even when experienced by preschoolers.

*Patricia J. Bauer, Marina Larkina, and Jacqueline S. Leventon*

See also Concepts, Development of; Memory, Neural Basis; Representations, Development of

**Further Readings**


**Exercise and the Brain**

Much has been written over the ages about the benefits of exercise. For example, Marcus Tullius Cicero stated that “it is exercise alone that supports the spirits, and keeps the mind in vigor,” and John Adams, the second president of the United States, suggested that “exercise invigorates, and enlivens all the faculties of body and mind. . . . It spreads a gladness and satisfaction over our minds and qualifies us for every sort of business, and every sort of pleasure” (de Mooy, 2003, p. 46). This review moves beyond opinions and conjecture to examine the scientific findings of exercise on brain and cognitive function.

**Animal Studies**

An emerging literature has documented the benefits of physical activity on brain and cognitive function. This research has been prompted, in part, by animal studies, which report that running influences brain function on at least three different levels. First, recent studies have shown that wheel running accelerates learning rates. In tasks such as the Morris water maze and other hippocampal-dependent spatial tasks, exercising rodents perform better than their sedentary counterparts. Second, exercising animals show robust differences in brain morphology.
Exercise and the Brain

compared to sedentary controls. Exercise induces neurogenesis, or the proliferation and survival of new neurons, and angiogenesis, or the proliferation of new capillaries. Although the functional significance of neurogenesis remains controversial, behavioral performance improvements associated with exercise suggest that newborn cells might facilitate learning and memory. Finally, exercise enhances levels of important molecules in the brain, including neurotransmitters and neurotrophic factors. For example, exercise increases the production and secretion of molecules promoting the formation of new blood vessels such as insulin-like growth factor and vascular endothelial growth factor. Brain-derived neurotrophic factor is also upregulated with exercise treatments and is necessary for long-term potentiation, a neural analog of long-term memory formation, and for the growth and survival of new neurons.

Human Studies

In humans, the effect of exercise on brain and cognitive function tells a similar story. Three types of experimental designs have been used to test this relationship: prospective-epidemiological, cross-sectional, and randomized trials. A review of the epidemiological literature suggests a significant relationship between physical activity and cognitive function in late adulthood. In a typical study, men and women are asked to report the number of times per week that they perform different aerobic, or physical, activities. These individuals are then followed for 5 to 10 years and the risk for disease is examined in relation to the frequency of self-reported physical activities. These studies have found that the incidence rate for Alzheimer’s disease and other neurological diseases is significantly higher for individuals who are infrequent exercisers compared to those who exercise more frequently.

Epidemiological studies have provided intriguing support for the relationship between physical activity and neurocognitive function. However, such studies cannot establish causal links between these constructs. Over the past several years, there have been a relatively small but increasing number of randomized trials in which relatively sedentary individuals, often over the age of 60, are randomized to an aerobic training group (i.e., walking, swimming, bicycling) and a control group that often entails nonaerobic activity such as toning and stretching. Training is usually conducted for an hour a day for several days a week and can last several months. Cognition, and sometimes brain function and structure, is examined prior to and subsequent to the intervention.

The results from randomized trials generally find that people in the exercise groups show improvements in cognitive function, while those in the control groups show either stability in cognitive function or even a slight decline. These studies also report that cognitive functions that are supported by the prefrontal and parietal cortex are enhanced more than cognitive functions that are less reliant on these brain regions. Empirical studies, meta-analyses, and reviews have reported that exercise regimens improve cognitive function across a number of domains and populations, including older adults and children, and in individuals with dementia and depression.

Some studies have addressed whether exercise training can positively influence the human brain using neuroimaging techniques. A number of cross-sectional studies have now reported that high-fit individuals have greater gray matter volume in the prefrontal, parietal, and temporal regions and greater white matter volume in the genu of the corpus callosum than their less fit counterparts. The hippocampus, a brain region involved in memory formation, shows a rapid rate of deterioration in individuals with Alzheimer’s disease, but it is larger in people who are more aerobically fit. Higher fitness levels are also directly related to better memory function that is supported by the hippocampus. Recent studies have also demonstrated that these effects generalize to individuals with dementia. That is, higher aerobic fitness levels are associated with the preservation of brain tissue in individuals with dementia.

Fewer neuroimaging studies have been conducted in randomized trials of exercise, but the few that have been conducted have shown that 6 months of exercise is enough to reverse age-related cortical decay in people over the age of 60. In at least one study, aerobic training (walking) increased gray matter volume in the frontal and superior temporal lobe and increased white matter volume in the genu of the corpus callosum, while the control group underwent a slight decline in cortical volume. These results suggest that even relatively short exercise interventions
Experimental Philosophy can begin to restore some of the losses in brain volume associated with normal aging.

In addition to changes that occur in the morphology of the brain, functional changes also occur in specific brain regions. For example, studies using functional magnetic resonance imaging have found that aerobically trained older adults, but not controls, show increased neural activity in the frontal and parietal regions of the brain and a reduction in activity of the anterior cingulate cortex, an area sensitive to behavioral conflict. Similar neurophysiological changes have been observed in individuals with multiple sclerosis and in the hippocampus of middle-aged adults. More recently, research has demonstrated that academic achievement scores and brain activity tend to be higher in more physically active children.

In sum, research on exercise and the brain has convincingly demonstrated a protective and profitable effect of physical activity on brain morphology and function.

Kirk Erickson

See also Aging, Memory, and Information Processing Speed; Memory, Neural Basis; Meta-Analysis; Working Memory

Further Readings


**EXPERIMENTAL PHILOSOPHY**

Experimental philosophy is part of a growing trend in philosophy, whereby scientific methods—formal and empirical—are brought to bear on questions of philosophical relevance. This entry explains why experimental philosophers have often focused on intuitions. It then examines the principal methods used by experimental philosophers. Finally, it illustrates the two main research programs in experimental philosophy with various examples.

Experimental Philosophy and Intuitions

Experimental philosophers typically rely on the methods of experimental psychology to study people’s intuitions—which, by stipulation, they take to be relatively fast, automatic, nonreflective judgments that apply concepts to particular objects, actions, events, or situations—about philosophical issues (causation, reference, free will, consciousness, right, what is permissible, etc.). Intuitions are used in various ways by philosophers, including as support for conceptual analysis, as claims about common sense, and as premises of arguments. The methods of experimental philosophers can be brought to bear on all these uses.

While experimental philosophers are typically concerned with understanding the psychological sources of intuitions (What mechanisms generate some specific intuition?), whether a given intuition is widespread (Does everybody have a given intuition?), and whether an intuition-producing mechanism is reliable (Does it deliver accurate information that can be used in philosophical theorizing?), they may also study behaviors, provided that these are of philosophical relevance. For instance, in a 2009 attempt to examine whether philosophical training in ethics promotes moral behavior, Eric Schwitzgebel has shown that ethics books are more likely to be stolen from university libraries than comparable books in other fields of philosophy.

The Methods of Experimental Philosophy

Experimental philosophers usually begin by constructing one or several vignettes (hypothetical stories). They then present the vignette(s) to participants, analyze the data, and consider what the judgments made by the participants reveal about the intuitions under investigation. The use of vignettes is not the only way to study intuitions, however. Adam Arico, Brian Fiala, Robert Goldberg, and Shaun Nichols have measured reaction times to show that people use low-level cues to classify an
entity as an agent and attribute conscious mental states to the entity. Examining what distinguishes philosophers’ skills from nonphilosophers’ skills, in 2010 Jonathan Livengood, Justin Sytsma, Adam Feltz, Richard Scheines, and Edouard Machery relied on a social-psychological questionnaire (the Cognitive Reflection Test) to highlight an aspect of philosophers’ temperaments: These tend to be more reflective than equally educated nonphilosophers. The work of psychologists such as Mark Alicke, Fiery Cushman, Joshua Greene, Tania Lombrozo, Jen Wright, and Liane Young is also closely related to the interests of experimental philosophers; these psychologists routinely use methods that go beyond recording the judgments elicited by vignettes. For instance, Greene and colleagues have used brain imagery and cognitive load methods (in which people are asked to complete a task while doing another distracting task) to examine the nature of the processes leading to moral judgments.

The Goals of Experimental Philosophy

Although experimental philosophers share much in common, there are some important differences among them regarding the role that intuitions should play in philosophy. Roughly, revisionists do not challenge the use of intuitions in philosophy but hold that experimental methods can make philosophers’ use of intuitions more reliable, while eliminativists are skeptical of the role of at least some types of intuitions in philosophy. The entry will discuss some prominent examples of each approach in turn.

There are several types of revisionist projects. Some revisionists uncover what intuitions people actually have and use these intuitions to make philosophical arguments. For example, philosophers commonly hold that phenomenal consciousness is an obvious aspect of our mental lives. If this were true, then people would distinguish those mental states that philosophers think have phenomenal properties (pain, olfactory and auditory experience, colors, etc.) from those mental states that are thought not to have such properties (e.g., beliefs). However, in 2010, Justin Sytsma and Edouard Machery provided some evidence that people do not do this; rather, people distinguish different kinds of mental states depending on the extent to which these mental states (seeing a color, smelling an odor, having an emotion, feeling pain, etc.) are valenced (i.e., have an affective component). Philosophers and psychologists also often hold that the function of folk psychology (the capacity to ascribe mental states to oneself and to others) is to predict and to explain behavior (particularly, others’ behavior). If this were true, then the ascription of mental states to an agent should not be influenced by the moral significance of her behavior. However, in 2006, Joshua Knobe provided a wealth of evidence suggesting that the attribution of mental states to an agent on the basis of her behavior is influenced by the moral nature of her behavior: Different mental states are ascribed depending on the moral significance of the behavior. For instance, when judging whether someone intentionally brought about a foreseen side effect, people judge that harmful, but not helpful, side effects are intentional. It would thus seem that folk psychology is essentially tied to our moral sense.

Finally, incompatibilist philosophers, who hold that free will determinism and determinism are incompatible, have argued that compatibilism (according to which free will and determinism are compatible) is a counterintuitive position. However, in 2006, Eddy Nahmias, Steve Morris, Thomas Nadelhoffer, and Jason Turner provided evidence that ordinary people do not think that free will is incompatible with determinism.

Other revisionists are concerned with how intuitions are generated, often to argue that some intuitions (but not others) are produced by unreliable mechanisms. Shaun Nichols and Joshua Knobe’s 2007 results suggest that people tend to have compatibilist intuitions when thinking in concrete, emotional terms but that they tend to have incompatibilist intuitions when thinking in abstract, cognitive terms. They then argue that the compatibilist intuitions should be discarded on the grounds that they are biased by emotions.

Eliminativists are concerned with uncovering a variety of unwanted psychological influences on some types of intuitions, and they use their empirical findings to argue that these intuitions cannot serve as reliable evidence in philosophical theorizing. In 2004, Edouard Machery, Ron Mallon, Shaun Nichols, and Steve Stich showed, for example, that intuitions about the reference of proper names vary across cultures: East Asians are more likely to have descriptivist intuitions (they judge that a proper name refers to the individual that satisfies the description associated with the proper name), while
Westerners are more likely to have causal-historical intuitions (they judge that a proper name refers to the individual that is historically linked to current uses of this name, whether or not this individual satisfies the description associated with the proper name). Since intuitions about reference appear to be influenced by a factor that arguably should not matter (viz., culture), and since there is no principled way to privilege certain intuitions, one might reasonably doubt that intuitions about reference should have a role in the philosophy of language. Another way to assess the reliability of intuitions is to see whether they are subject to context and ordering effects. In 2008, Stacey Swain, Joshua Alexander, and Jonathan Weinberg thus provided some evidence that people’s disposition to ascribe knowledge sometimes varies depending on whether the situation is contrasted to a clear case of knowledge or whether it is contrasted to a clear case of ignorance. Thus, widespread disagreement and context, ordering, and demographic effects can all be used by the eliminativist to show that certain intuitions are unreliable and should thus be eliminated from philosophical practice.

Edouard Machery and David Rose

See also Access Consciousness; Concepts, Philosophical Issues; Emotion and Moral Judgment; Folk Psychology; Freedom of Action

Further Readings


Explanation of Action

Certain explanations are characteristic of purposive action. The question, “Why did Ayesha poke the chimpanzee?” could be answered by citing Ayesha’s desire to retrieve her glasses and her belief that poking the chimpanzee would make it return them to her. Explanations such as this one seem to be characteristic of actions such as Ayesha’s. This entry pursues two questions. First, which features distinguish explanations characteristic of action? Second, what is the nature of such explanations—are they causal, are they teleological, and are they instrumental?

The term action can be used in a broad sense or a narrow sense. In the broad sense, action encompasses reflex actions, motor actions, arational actions (such as jumping for joy), and more. A narrow sense of action restricts the term to purposive actions—that is, to actions that are attempts, successful or unsuccessful, by an agent to achieve some end of theirs. In this entry, the term action is used in this narrow sense. A paradigm case of action in the narrow sense is Ayesha’s poking the chimpanzee to retrieve her glasses from it.

Which Features Distinguish Explanations Characteristic of Action?

An explanation characteristic of action is one with a certain combination of features that is only found in explanations of why actions happened and not in explanations of why other types of events happened. Which features make up this combination? One is an appeal to reasons that may rationalize the action. Another feature of explanations characteristic of
action is arguably that they invoke psychological states such as beliefs and desires.

**Reasons**

Explaining why any event occurred involves giving reasons that explain why the event happened. We can answer the question, “Why did the apple fall?” by giving a reason, as in “Because it was ripe.” That the apple is ripe is a reason that explains why it fell. A distinguishing feature of explanations characteristic of action is that they involve reasons that not only cause but may also rationalize the action.

Philosophers distinguish between two notions of reason for action, normative reasons, and operative reasons. To illustrate, consider the proposition that poking the chimpanzee may cause it to return Ayesha’s glasses. Whether this is an operative reason depends on whether this proposition plays any role in explaining why Ayesha did in fact poke the chimpanzee. By contrast, whether it is a normative reason for Ayesha to poke the chimpanzee depends on whether this proposition has the potential to rationalize her action—that is, on whether the truth of this proposition would support the conclusion that it was rational for Ayesha to poke the chimpanzee given all the reasons she has for and against so acting. Normative reasons are not invariably operative reasons; sometimes propositions that would rationalize an action are not relevant to explaining why the agent acted, and conversely, operative reasons are sometimes not normative reasons.

That operative reasons are sometimes also normative reasons distinguishes explanations characteristic of action. Explanations of events other than actions involve reasons that are not even potentially normative reasons for those events (i.e., the ripeness of the apple is not even potentially a normative reason for its fall).

Where operative reasons are also normative reasons, it is plausible that their explanatory role depends in some way on their normative role. Contrast the following explanation with that offered at the start of this entry:

Why did Ayesha poke the chimpanzee? Because she believed it might bite her, and she wanted a drink.

The explanation offered at the start of this entry stands alone—although not a complete explanation, it could serve as an adequate answer to the why question all by itself. But the explanation above, even if factually accurate, could not be an adequate answer to the why question all by itself. In general, where a partial explanation provides operative reasons that are not normative reasons, merely citing those reasons does not produce an explanation that stands alone. This suggests the role of reasons in explaining action may be linked to their rationalizing role.

**Beliefs**

When explaining an action, it is common to include falsehoods as operative reasons. Operative reasons for Ayesha’s poking the chimpanzee include the proposition that poking it will cause it to return her glasses. As it happens, this is false. No amount of poking would cause this particularly stubborn chimp to cooperate. But how can a falsehood explain anything? In general, it is not possible to explain why an event happens by citing falsehoods. If an apple is not ripe, you cannot explain why the apple fell by appeal to the proposition that it is ripe. Why are explanations of action apparently different in this respect?

The leading view is that operative reasons explain an action only by virtue of standing in some relation to the contents of the agent’s beliefs or other psychological states. When we explain why an apple fell by giving a reason (e.g., that it is ripe), the reason is explanatory by virtue of being a fact. But when we explain why an action occurred by giving an operative reason, the reason explains by virtue of being the content of a belief. Ayesha poked the chimpanzee not because poking it would in fact cause it to return her glasses but because she believed this to be so. This is how a proposition can be an operative reason for her action despite its falsity.

**Desires and Other Pro-Attitudes**

Many actions are in part a consequence of the agent’s desires. Ayesha pokes the chimpanzee only because she wants her glasses back. Had she wanted the chimpanzee to retain them, she wouldn’t have poked it. But apparently not all actions involve desires. It is natural to think agents sometimes act out of pride or conviction as opposed to desire. The notion of a pro-attitude enables us to avoid the connotations “desire” carries. A pro-attitude is an attitude toward an action that, like a desire, disposes an
agent to act in certain ways under particular conditions. Examples other than desire include pride and intention as well as attitudes related to convictions and values.

What role might citing pro-attitudes play in explaining actions? We saw above that explanations characteristic of action sometimes rationalize actions in explaining why they occur. Now actions, like all events, fall under more than one description. Ayesha’s poking the chimpanzee is Ayesha attempting to retrieve her glasses, and this event is identical to Ayesha’s getting herself thrown out of the zoo. What is rationalized is not the bare event but rather the event-under-a-description. To rationalize Ayesha’s attempting to retrieve her glasses is not necessarily also to rationalize her getting herself thrown out of the zoo. One function of explanations characteristic of action is to reveal descriptions under which it is supposed to be rational for the agent to have performed this action. This can be achieved by citing a desire (or pro-attitude) that the agent has about a type of action. In the explanation given at the start of this entry, Ayesha desires to retrieve her glasses. The poking action she performs falls under this description—or so she believes—and is therefore rational from her point of view, in at least one sense of rational.

In a widely held view that has generated intense discussion among philosophers, action explanations canonically involve a belief-desire pair that jointly rationalize the action. The desire (or pro-attitude) is to perform a type of action and the belief entails that the action the agent performs is an action of the desired type. The explanation given at the start of this entry has this canonical form. Part of the point of citing the agent’s desire is to identify a description under which the action is rational from the agent’s point of view.

What Is the Nature of Action Explanation?
The distinctive features of action explanation described above lead to questions about the relation between explanations of action and explanations of other types of events. It is widely held that explanations of action are causal explanations, and some claim that explanations of action are also teleological. A further view is that explanations of action are instrumental.

Causal
Explanations characteristic of action are explanations of why an event happened. This is grounds for holding that such explanations are causal explanations.

Causation has a special role to play in explanations of action. Agents sometimes have reasons that would rationalize actions they perform but do not explain why they act (in terminology introduced earlier, these are normative but not operative reasons). To illustrate, Ayesha wants Tim to think she is brave and knows that poking a chimpanzee will achieve this. As it turned out, she did poke a chimpanzee but not for this reason. The desire to be thought of as brave played no role at all on this occasion; her sole motive was to retrieve her glasses. When an agent has a reason that rationalizes an action she performs, what more is involved in this reason’s being a reason that explains why she acts? The standard answer is that it is necessary (but not sufficient) for the reason to cause her action. Or more accurately, the psychological state or states that constitute her having that reason—the relevant beliefs and desires—must be among the causes of her action.

Some philosophers deny that action explanation is causal. One challenge for them is to provide an alternative account for the difference between an agent’s having a reason that merely rationalizes some action she performed and a reason that explains why the agent acted.

Teleological
Teleological explanations explain why events occur by appealing to their outcomes. At one time teleological explanations were regarded with suspicion. They appear to conflict with the principle that causes precede their effects, for they cite an outcome to explain why an event occurred. Philosophical accounts of teleological explanation have since shown that such explanations can be legitimate. For an illustration, consider the proposition that Atta ants cut leaves to fertilize their fungus crops. This proposition teleologically explains why Atta ants cut leaves, by appeal to fertilizing, which is an outcome of their cutting. How can the outcomes of events explain why they occur? On Larry Wright’s analysis of teleological explanation, the explanation amounts to this: (a) Cutting leaves tends to
Explanatory Gap

The term *explanatory gap* was first introduced by Joseph Levine in 1983, though the problem to which the term refers is quite old. Consider a quite
ordinary conscious experience, such as seeing the color of a leaf in the fall. We know quite a bit about the causal mechanisms responsible for the experience, starting from the light being reflected from the leaf, the light then hitting the retina, and then the subsequent neural activity in the visual areas of the brain. The problem is that when one focuses on just what it's like to view the bright orange color of the leaf, it doesn't seem as if the information about the neural activity in the visual system at that moment adequately explains why the leaf looks just the way it does. This gap in our understanding about how the neural mechanisms in the visual system result in the conscious visual experience of the leaf’s color is the explanatory gap.

**Functional Versus Intrinsic Properties**

To appreciate the force of the problem, it’s necessary to distinguish between a mental state’s functional properties and its intrinsic properties. The functional properties of a mental state, such as a visual experience, are determined by the role that that state plays in the overall functioning of the subject. So, for example, visual experiences of color carry information concerning the surface properties of the objects viewed and also contribute to determining the behavior of the subject in the relevant circumstances. By distinguishing red from green surfaces, we often thereby distinguish ripe from unripe fruit, which helps us decide whether it's advisable to eat what's in front of us.

Notice that a functional role is a rather abstract characterization that tells us, as it were, what job is done, but not how it's done, or what precisely it is that’s doing it. We identify how a role is carried out by appeal to a state’s intrinsic properties. In the case of the visual system, the bio-chemical properties of the neural circuits that fire in response to light stimuli explain how discriminations between red and green surfaces, for example, are made. This is similar to the way that descriptions of the hardware circuitry of a computer explain how the functional roles defined by the programs it’s running are carried out.

Now we can see the relevance of the distinction for the explanatory gap. Consider again the qualitative character of an experience of viewing a bright orange leaf in the fall. The question is, what explains this particular qualitative character? Why is it like this and not some other way, to see the leaf? The obvious place to look for an answer is to the scientific story about the mechanisms from the retina through the visual system, how each neuron responds to the input from its neighbor upstream and transmits the relevant impulse to its neighbor downstream. But here is the problem. It seems that the most we can get from an account of the neural mechanisms in the visual system is an explanation of how the functional role played by a visual experience is carried out. In other words, a description of the physical properties of the neural mechanisms explains how they are caused to operate as they do and how their operations cause the downstream effects on cognition and behavior. It is unclear how knowledge of the physical mechanisms in the visual system can explain anything other than how they are caused and what they cause in turn.

However, when we consider the qualitative character of an experience, such as seeing a bright orange fall leaf, it certainly doesn’t seem that we can capture what it is like purely in terms of the causal, or functional, role played by the experience. There seems to be an intrinsic, nonfunctional character that we want explained. One way to see this is through the inverted spectrum hypothesis. The idea here is to imagine a creature whose color experiences were the complements of those of normal perceivers. They had greenish experiences where others had reddish ones, and bluish experiences where others had yellowish ones. Because all the relations among the color experiences would be maintained, the functional, or causal, roles of their experiences would match normal perceivers. They would call ripe fruit “red” and unripe fruit “green,” just like the rest of us, only the quality of their experiences would be different. The problem, then, is how to explain why our experiences are reddish where theirs are greenish in terms of the differences in our neural mechanisms. What makes one neural mechanism constitute a reddish experience and another a greenish one? Since the difference in the intrinsic character of the experiences isn’t a difference in their causes and effects, there is no way for the descriptions of the underlying neural mechanisms to explain the difference. That is, there is an explanatory gap between the physical level and the level of conscious experience.

**The Problem of “Alien” Minds**

The import of the explanatory gap is evident in the following conundrum as well. Suppose we encounter
alien life that is constituted by very different physical materials than we are but nonetheless manifests the same functional profile, just like two computers with different hardware are able to run the same programs. Now imagine we and the aliens are looking at the same bright orange leaf. After learning our language, the alien can describe it as orange. But is the alien having an experience just like ours? How would we know? What information could decide this question? Now, if we could explain the qualitative character of our experience in terms of the relevant underlying physical mechanisms, we would be able to tell, since we could look for just those features in the alien that are responsible for the quality of our experience in us. However, since we don’t know what it is about our visual system that explains the quality of our visual experience, we don’t know what to say about the alien. We know the alien can make the same discriminations as us. What we don’t know is what it is like for the alien, what the quality of the alien’s experience is.

Consequences for Materialism

Materialism is the doctrine that mental states and properties are, at bottom, physical states, and properties. The explanatory gap is a problem for materialism because it casts doubt on the claim that conscious experiences are ultimately physical states. Materialists have responded to the explanatory gap in two ways. First, some have argued that in fact the qualitative character of an experience is reducible to some aspect of its functional role—many say it’s a matter of the visual state’s informational content. If this reduction to functional role goes through, then we can close the gap since we have already seen that descriptions of underlying physical mechanisms are well suited to explain how functional roles are carried out or realized.

The second materialist strategy for addressing the explanatory gap is to concede its existence but argue that it stems from a feature of our cognitive architecture that is fully consistent with materialism. The idea is that we are cognitively structured in a way that we cannot integrate information from the third-person perspective from which scientific descriptions of our mental activity are derived with information we get in the first-person mode. So when I consider the character of my experience as I experience it, the representations involved cannot be connected to those I obtain from a scientific description in a way that allows me to see how they amount to the same thing.

So far the debate over materialism and the explanatory gap continues without any philosophical consensus over the best response, whether it be one of the two materialist strategies mentioned or the rejection of materialism about conscious experience altogether.

*Joseph Levine*

**See also** Access Consciousness; Consciousness and Embodiment; Emergence; Mind-Body Problem; Neural Correlates of Consciousness; Physicalism; Realism and Instrumentalism; Reductive Physicalism

**Further Readings**


**EXTENDED MIND**

The extended mind (EM) is a radical thesis about the constitution of minds. Minds are constituted by neural, bodily, and environmental states and processes. EM theorists believe that minds are hybrid entities assembled from the continuous and dense interactions between brains, bodies, and the local environment. The theory has been developed by cognitive scientists from a wide variety of disciplinary backgrounds, including philosophy, psychology, and cognitive anthropology. The theory was given its name in a famous paper by Andy Clark and David Chalmers and this has been taken to be the core theoretical statement. However, the theory was also developed in slightly different ways by Ed Hutchins as distributed cognition, Susan Hurley as vehicle externalism, Robert Wilson as wide computationalism, Mark Rowlands as environmentalism, and Richard Menary as cognitive integration (some of these will be described below). The first section
will outline the original thesis as presented by Clark and Chalmers in their article. The second section will provide some of the standard criticisms that the position has attracted, and the final section will outline some of the alternative ways that the thesis has been developed by others.

**The Original Thesis**

Clark and Chalmers begin their article by articulating an active externalism. This is a novel form of externalism about the mind; traditional externalism claims that the contents, or meanings, of mental states such as beliefs are determined by the nature of the external world. Active externalism concerns the role of the environment in driving our cognitive processes. Active externalism is committed to the idea that some cognitive processing is constituted (in part) by active features of the environment. The idea can be illustrated by the use of David Kirsch and Paul Maglio’s concept of an epistemic action. Pragmatic actions are aimed at achieving a noncognitive goal, such as a full stomach. Epistemic actions are aimed at achieving a cognitive goal such as solving a problem by directly manipulating the environment. For example, experienced Tetris players rotate “zoids” as soon as they appear to determine where they fit best. The rotations are not directly tied to the goal but alter the task space so as to make reaching the goal easier, or more achievable. Similarly, when experimental subjects are allowed to physically rearrange tiles in a game of scrabble they create more words than those who are not. The EM theorist takes examples of epistemic action to be cases of active externalism. Why though should we draw this conclusion? Kirsch and Maglio’s research shows that our problem-solving routines are often distributed, or extended, across brain, body, and environment. If we think of the problem-solving space as a state space (a state space can be thought of as a finite number of state variables, which define the state of the system at a particular time), then bodily “internal” organizing processes (i.e., neural ones) and bodily “external” organizing processes (i.e., physical manipulations of the environment) are all states in the same problem-solving space. To understand this better, we need to look at the idea of coupled processes.

A coupled process is one where there is a reciprocal causal interaction between two states. The coupled process is part of a system when the interaction between the states in the coupled process controls some behavior of the system. These kinds of coupled processes in dynamical systems can be found throughout the natural world but also in artifacts such as the Watt flywheel governor. It is now easy to see that the organizing processes that involve “internal” states and “external” states constitute a coupled process that controls problem-solving behavior of a cognitive system. It is important to note that EM theorists are not committed to the view that artifacts and tools are themselves cognitive or mental, nor that a simple causal interaction between two states X and Y makes X part of Y; these are mistaken caricatures of the position.

The article also makes use of a principle of parity to identify external processes that might qualify as cognitive processes. The parity principle tries to motivate the idea that the location of a process should not, by itself, discount a process from being cognitive. The parity principle can most clearly be stated as the following: *If an external process leads directly to the completion of a cognitive task, and if we would call the external process cognitive if it were an internal process, then the external process is part of our cognitive processing of the task.* This is because the only relevant difference is that the process does not occur in the brain. If we claim that the only factor by which we deny the process cognitive status is its nonneural nature then we commit some form of neural chauvinism.

This principle is in tension with the account of coupled processes that has been given above, because in that account coupled processes involve interacting states that are both bodily internal and bodily external, whereas the parity principle focuses on bodily external processes without putting them in the interactional setting of a coupled process.

An example of the parity principle at work is the now famous Otto thought experiment, introduced in Clark and Chalmers’s article. Consider Inga: she hears that there is a Rothko exhibition on at the Museum of Modern Art and she decides to go to it. Inga recalls the location of the Museum of Modern Art from a long-standing biological memory, which causes her to go to 53rd street. Now consider Otto: He has a mild form of Alzheimer’s and carries a notebook for the retrieval of information. He has all sorts of useful information about places and people, addresses, and names, and so forth. Otto takes his
notebook with him wherever he goes and refers to it frequently. On hearing about the same exhibition as Inga, he decides to go, but Otto retrieves information from his notebook concerning the location of MOMA. This causes him to go to 53rd street.

The EM theorist holds that Otto’s notebook plays the role of a dispositional memory in Inga (a memory that is available to consciousness but can be acted upon without being consciously accessed), and as such the two cases are on a par. We should count the process of Otto’s retrieving information from his notebook as a cognitive process (or part of a coupled process) even though that process is not located in his brain. This is the case only if Otto’s notebook plays the same role for Otto that biological memory plays for Inga. We might be inclined to think that the information in Otto’s notebook is reliably available to him and guides his actions in just the sort of way that beliefs are usually supposed to. The information is available and functions just like the information that constitutes nonoccurrent beliefs (occurrent beliefs are those of which we are currently aware); the only difference is the location of the information. Taken in this way, EM turns out to be a species of functionalism: It is function, not physical constitution, that matters to mentality. As long as information plays the relevant role (or function) it is a belief regardless of location. Hence, the mind is extended into the world.

**Criticisms and Responses**

There are a number of criticisms that have been leveled at EM; the two most prominent are the causal coupling fallacy and the mark of the mental.

**The Causal Coupling Fallacy**

The objection was first stated by Fred Adams and Ken Aizawa: When some object or process is coupled to a cognitive agent in some way, the EM theorist slides to the conclusion that the object or process constitutes part of the agent’s cognitive apparatus. Schematically the fallacy looks like this: Just because X is causally related to Y, it does not follow that X is a part of Y. Given the definition of a coupled process above, it is hard to see how EM theorists are guilty of such a simple fallacy. For the reciprocal causal interaction of X and Y is part of a wider system Z. We identify X and Y as parts of Z because of their contribution to Z’s behavior. For example if we damage or remove X or Y, Z’s behavior will change. Take Otto. Otto’s memory system is constituted by the fragmentary processes in his brain and the manipulation of written vehicles in his notebook. If we remove Otto’s notebook we will have changed the behavior of his memory system: In fact, he will be unable to recall where MOMA is located. Adams and Aizawa’s criticism could still be reformulated in a way that is damaging to the EM theorist, because the EM theorist appears to think that the coupling between X and Y make them a part of Z, so the fallacy would be that a coupled process involving a reciprocal interaction between X and Y could not constitute (even in part) a wider system.

However, coupled processes such as this often become constituents of a wider system. Take intestinal flora for example. We acquire them after birth and they have an immediate and ongoing effect on the behavior of our digestive system. Indeed, without the constant dynamic interaction between intestinal flora and our digestive organs we would not be able to digest certain kinds of carbohydrates. According to Adams and Aizawa’s fallacy, intestinal flora should not be counted as part of our digestive system.

**The Mark of the Mental**

Is there a mark of the cognitive or the mental that all cognitive and mental states must have to be considered cognitive or mental? Such a mark if it could be determined, would allow us to determine whether the inscriptions in Otto’s notebook could really be part of his memory system. Adams and Aizawa stipulate that the mark of the cognitive is representations with underived content (or meaning); they are very careful to say that a cognitive process must involve at least some underived content. This means content that is not derived from another mental or cognitive state. Examples of representations with derived content are external representations such as written sentences or road signs which get their contents from mental or cognitive states. In other words, for a written sentence or a stop sign to mean anything there need to be minds that interpret the sentence or the stop sign to mean what they do. By contrast, your thought—that it is a sunny day today—is not derived from some other mind; that thought is underived. It then becomes clear that the written sentences in Otto’s notebook have
only derived content and therefore lack the mark of the cognitive. There are several things to note about this objection. The first is that it is unusual for a science to have articulated a mark of its subject matter; it is unclear exactly what the mark of the physical or the biological might be, and physicists and biologists have been doing very well without one. The second thing to note is that as a stipulation the proposed mark has no independent motivation and such a motivation is required. The final thing to note, before looking at a response to the objection, is that the concepts of representation and content are highly contested, and there is no real agreement on what the criteria are for representation, nor for how a representation gets its content determined.

There is an easy response to the mark of the mental objection, which is to accept the stipulation and allow that in a coupled system such as Otto’s there are both representations with underived content and representations with derived content. Therefore, as long as a coupled process involves at least some underived content, Adams and Aizawa’s objection is met. A more restrictive application of the objection would be that only a process involving representations with underived content can count as cognitive.

**Other Formulations**

Other formulations of EM center on the concepts of complementarity, integration, and manipulation. Rowlands develops his version of EM by proposing a manipulation thesis in which some cognitive processes involve the manipulation of external information-bearing structures. The manipulation thesis is entirely consistent with EM as developed by Clark and Chalmers; however, it does not depend upon the parity principle as a motivational tool. Menary takes the key to EM to be a matter of the integration of internal and external manipulations, which is also consistent with EM as presented above. Sutton has stressed the complementary roles of internal and external resources drawing on the work of Merlin Donald.

*Richard Menary*

**See also** Anti-Individualism About Cognition; Distributed Cognition; Mind-Body Problem

**Further Readings**


**EYE MOVEMENTS DURING READING**

This entry briefly details what is known about eye movements during reading. It includes an overview of properties of the eye, types of eye movements, commonly used eye movement measures, research methodologies, a discussion of visual and linguistic factors of words that affect eye movements during reading, and a brief mention of a computational model that attempts to simulate these influences.

**Eye Movements and Properties of the Eye**

When we read, our eyes don’t glide smoothly across the page. Rather, we make a series of eye movements (called saccades) separated by fixations. Eye movements are tightly linked to cognitive processing and offer the opportunity to rigorously test hypotheses about how the mind operates during reading. The reason we move our eyes is so that we can place the fovea—that part of our retina with the highest acuity (resolution)—over the text we wish to process. The fovea extends 2 degrees of visual angle around fixation and outside of this area acuity drops off rapidly in the parafovea (2–5 degrees) and periphery (5 + degrees). Saccades, which typically last 25 to 60 milliseconds (ms) (depending on their size), are considered to be ballistic—so that it is difficult to change trajectory once initiated—and take approximately 175 ms to plan and execute if no other cognitive processing must be completed. Between the saccades, our eyes are in relatively stable states— fixations—which last approximately 225 to 250 ms. Information is obtained during
fixations as we are essentially “blind” during saccades because of saccadic suppression. For the most part, overt attention (where the eyes fixate) and covert attention (where the mind attends) are tightly linked in reading.

In reading, the average saccade extends 7 to 9 letter spaces. Letter spaces are the appropriate indicator of how far the eyes move since regardless of the reading distance of the text (which modulates text size on the retina), eye movements extend the same number of letter spaces, not the same degree of visual angle. Not all saccades move forward in the text; 10% to 15% of saccades are regressions, which means they return to previously read or skipped text. More difficult text leads to more regressions as well as shorter saccades and longer fixations.

**Eye Movement Measures**

Although the abovementioned metrics—global measures of processing—are important, researchers more often discuss local measures of processing, such as *first fixation duration* (duration of the first fixation on a word), *single fixation duration* (duration of the fixation on a word when the word was only fixated once, excluding regressive fixations), and *gaze duration* (the sum of all fixations on a word before the eyes move to another word). These measures are more useful because they are more robust to skipping and regressions, whereas global measures have the implicit assumption that all words are fixated once. Approximately a third (31%) of words in text are skipped—not directly fixated. In general, these words are short, high-frequency words with little semantic content (e.g., *the, in,* and *and*) or words that are highly predictable from prior text. Conversely, words likely to be refixated (i.e., fixated more than once before the eyes continue) tend to be longer and less frequent. The abovementioned local measures are useful to estimate how long it takes to process a word.

**Properties of Words That Affect Eye Movements**

The amount of time spent fixating a word is influenced by many linguistic factors, such as frequency, predictability, lexical ambiguity, age of acquisition, familiarity, and so on. Even when a word disappears or is masked 50 to 60 ms after it is fixated, the time spent fixating on that word is highly influenced by frequency and is very similar to reading time when the word does not disappear. However, when the following word disappears at the same time as the fixated word, reading is greatly disrupted, indicating that the word to the right of fixation is important for normal reading.

**Computational Models of Eye Movements During Reading**

A number of computational models of eye movements in reading have been proposed over the past few years. Space does not permit a discussion of these models here, but a comprehensive review can be found in a special 2006 issue of *Cognitive Systems Research*. Most models can account for all the previously mentioned data, accommodating both global and local measures. One of the most influential and robust models is the E-Z Reader model, which can predict fixation durations, which words will be skipped, and which will be refixated.
**Summary**

Eye movements are an excellent tool to study the cognitive processes underlying reading, as they are determined by not only characteristics of the reader (e.g., reading skill) but characteristics of the text as well (e.g., text difficulty, word frequency, predictability). Eye movements have been rigorously studied over the past three or four decades, and with advancements in eye tracking technology this progress will most definitely continue.

*Keith Rayner and Elizabeth R. Schotter*

*See also* Dyslexia, Acquired; Word Recognition, Visual

**Further Readings**


---

**Eyewitness Memory**

The legal system relies heavily on the testimony of eyewitnesses at criminal and civil trials to establish facts about previous events. Because human perception and memory are fallible, scientific psychology has a longstanding interest in helping the legal system understand the conditions under which such testimony can be mistaken. Psychologists use theories of memory to articulate basic principles in how the mind acquires, stores, and retrieves information and to explain how problems at each of these levels can produce erroneous eyewitness accounts.

Writings on eyewitness psychology date back to the late 1800s in Europe. Hugo Munsterberg, often considered the father of applied psychology, popularized eyewitness psychology in the United States in the early 1900s. These early writings were mostly attempts to apply existing theories of perception and memory to courtroom testimony rather than systematic experiments directed at eyewitness situations. Beginning in the 1970s, scientific psychologists began to more programatically conduct novel experiments to study eyewitness testimony by creating filmed or live staged events (e.g., mock crimes) for unsuspecting people. Eyewitness psychologists attempt to experimentally isolate variables that increase and decrease rates of eyewitness error. The accumulated hundreds of such experiments in the published literature now constitute a foundation for expert testimony in civil and criminal trials and provide a basis for advising the legal system regarding how to best collect and preserve eyewitness evidence. This entry briefly overviews several major domains of eyewitness literature, including event testimony, identification testimony, and eyewitness confidence.

**Event Testimony**

One domain of particularly important work, pioneered by psychologist Elizabeth Loftus in the mid-1970s, concerns postevent information. Postevent information is “information” (either true or false) that an eyewitness might acquire after the event that is incorporated into later testimony as though it was part of what the witness had originally seen. One common example of postevent information is a misleading question. For instance, after seeing a man walk in through a front door and shoot a store clerk, an eyewitness might be asked if the gunman who came in through the back door said anything before he shot the clerk. When later asked to describe the actions of the gunman, the witness is more likely to say that he came in through the back door than if that question had not been asked. The phenomenon of people incorporating postevent information into their recollections of earlier events contrasts sharply with the legal system’s general presumption that although a person might forget details—and therefore would not report them later—their memory would not incorporate new details.

There remains some debate among psychologists regarding exactly how postevent information influences memory. The original memory could be actually replaced by the postevent detail, or it could be blended with the postevent detail, or an entirely new, separate memory could compete with the original memory for retrieval. There are experimental results consistent with each of these processes, and it is possible that any of the three can happen, depending on the circumstances. In any case, there are clear regularities to the conditions that increase and
decrease the magnitude of postevent information effects. For instance, postevent information effects are more pronounced when the original memory is weak rather than strong. Hence, it tends to be easier to find postevent information influences on peripheral details of an event than on central details of an event. Likewise, postevent information effects tend to be stronger when the witnessed event was in the more distant past. Furthermore, postevent information is more likely to be incorporated into memory testimony when the postevent information is plausible rather than implausible.

Identification Testimony

A considerable amount of the work on eyewitness testimony in recent years has been concerned with eyewitness identification testimony. Identification testimony refers to a claim by an eyewitness that a specific person was seen committing a particular act. In cases where the identified person was previously known to the eyewitness (e.g., a relative), accuracy of identification is not usually an issue. But in cases where the perpetrator is a stranger, the issue of mistaken identity is often a concern. Experiments on eyewitness identification from lineups following simulated crimes have illustrated that mistaken identification is not at all uncommon. Numerous factors have been shown to hamper the ability of eyewitnesses to accurately identify someone from a lineup. For instance, other-race identifications are less reliable than same-race identifications, stress during witnessing impairs eyewitness identification performance, changes to a perpetrator’s appearance (perhaps because of disguise) are highly problematic for later identification, and the use of a weapon tends to draw the witness’s attention away from a perpetrator’s face and toward the weapon.

Many eyewitness psychologists have focused their research specifically on system variables, which are the parts of the investigatory process affecting the accuracy of eyewitness identifications over which the justice system has control. For instance, instructions given to eyewitnesses prior to their viewing a lineup, the composition of a lineup, and the verbal and nonverbal behaviors of a lineup administrator can each influence both eyewitnesses’ identification decisions as well as the certainty that they express in their identification decision. A particularly difficult situation for eyewitnesses is when they view a lineup in which the actual perpetrator is not present. When the actual culprit is not in a lineup, eyewitnesses tend to select someone who looks the closest to their memory of the culprit than do the remaining lineup members. Based on this research, eyewitness scientists recommend that police conduct sequential lineups, which encourage eyewitnesses to make more absolute judgments about each lineup member. In a sequential lineup, eyewitnesses only see one lineup member at a time and must decide yes or no before moving on to the next lineup member. Because eyewitnesses do not know how many lineup members they will see, they cannot simply pick the closest—the next lineup member may be the actual perpetrator. This can reduce the likelihood of a mistaken identification when the perpetrator is not present, as witnesses would likely get through the entire lineup without identifying anyone. Another type of identification procedure is called a showup. Whereas a lineup presents the eyewitness with the suspected person embedded among known-innocent distracters, a showup presents the eyewitness with only one person (the suspected person). Showups are not considered to be good tests of eyewitnesses’ memories because they tend to be suggestive, and because identification errors cannot be distributed to known-innocent distracters.

Eyewitness Confidence

One of the central issues in the scientific literature on eyewitness testimony has been the confidence with which eyewitnesses testify about their memories. Trial simulation studies have shown that the confidence with which an eyewitness asserts some fact (e.g., “Number three in the lineup is definitely the person I saw” or “I am positive that the man came through the rear door”) is a primary reason why people will accept the statement as being accurate. However, witnesses can give highly confident testimony on matters for which they are mistaken. Dissociations between confidence and accuracy appear to occur in large part because some of the variables that affect confidence do not also affect accuracy. For instance, repeated recall of a mistaken memory tends to inflate confidence but does not enhance accuracy—like wearing a path into a patch of grass, repeated recall entrenches the memory so that the witness is confident while traversing the path, even though the path might be the wrong
direction to take. Also, confirming feedback for a mistaken memory (e.g., “Good, you identified the right person”) inflates confidence and leads witnesses to think that they were positive all along when, in fact, they had doubts when they first made their identification.

**Miscarriages of Justice Based on Mistaken Eyewitness Testimony**

Since the advent of forensic DNA testing in the 1990s, DNA testing has been used to assess the innocence claims of some convicted felons for whom DNA-rich, biological evidence had been collected and preserved after trial. This has resulted in over 230 exonerations (as of May 2009) based on DNA proof of actual innocence. Slightly over 75% of these convictions turn out to be cases of mistaken eyewitness identification testimony. These real-world, mistaken, eyewitness cases tend to have the characteristics that eyewitness researchers have described in their experimental studies, such as an overrepresentation of other-race identifications, highly confident trial testimony, suggestive procedures, and identifications from lineups that did not contain the actual perpetrator.

Almost all DNA exoneration cases are cases that involved sexual assault; these are among the only types of cases for which there is biological evidence that can provide definitive proof of innocence. Although 230 such cases is not a large number, few claims of innocence based on mistaken identification can actually be tested because, for the vast majority of cases, no DNA is available that could trump an eyewitness’s account. As a result, there is no way to know the actual number of wrongful convictions based on faulty eyewitness identifications. Nevertheless, because mistaken eyewitness identification accounts for more DNA exonerations of innocent people than all other causes combined, the American legal system has begun to show increased receptivity to what psychological science has to offer on the issue of eyewitness testimony. This includes an increasing tendency to permit expert testimony regarding eyewitness issues at trial, and an increasing tendency for law enforcement to follow recommendations from eyewitness scientists and adopt lineup procedures that lower the chances of mistaken identification.

_Gary L. Wells and Elizabeth A. Olson_

*See also* Emotion and Working Memory; Face Perception; Memory, Interference With; Memory, Neural Basis; Memory Recall, Dynamics

**Further Readings**


FACE PERCEPTION

As Humpty-Dumpty observed in Lewis Carroll’s (1871) classic, Through the Looking Glass and What Alice Found There, all faces share the same basic configuration. On encountering Alice, Humpty-Dumpty complained,

You’re so exactly like other people, . . . the two eyes, so (marking their places in the air with his thumb), nose in the middle, mouth under. It’s always the same. Now if you had the two eyes on the same side of the nose, for instance—or the mouth at the top—that would be some help. (p. 133)

Faces are also highly animated, with changing expressions and other movements resulting in a variety of views. This combination of structural similarity and dynamic variation creates challenges for visual analysis and recognition. Yet we readily extract a wealth of social information from the face, determining the individual’s identity, gender, age, attractiveness, emotional state, and direction of attention. We can also use remarkably subtle cues, such as the skeptical elevation of an eyebrow or an inviting glance, to adjust our social interactions. And, unlike Humpty Dumpty, we do all this rapidly and with little apparent effort. Over the last three decades researchers have discovered much about the cognitive and neural mechanisms that make all this possible. This entry reviews what is known about how we recognize faces and how this capacity develops during infancy and childhood. It also considers the perception of other aspects of faces, such as their attractiveness and emotional expressions. Finally, the entry examines the neural mechanisms underlying face perception, and how damage to these can result in acquired prosopagnosia.

Face Recognition

Humpty Dumpty’s complaint illustrates one of the central challenges in recognizing individual faces. Unlike many other visual objects, we do not distinguish faces by their parts alone or even by the basic arrangement of these parts. Rather, successful face recognition depends on sensitivity to subtle spatial relations between facial features, sometimes referred to as configural coding. Unlike most other objects (except perhaps words), faces are also coded holistically, with little explicit coding of component features. Turning faces upside down disrupts recognition of faces far more than other mono-oriented objects, and this disproportionately large effect is likely due to inversion dramatically reducing our sensitivity to spatial relations (see Figure 1) and disrupting our holistic coding of faces.

Some faces are easier to remember than others, and this may be a consequence of how the similarities and differences among faces are represented. Individual faces can be thought of as “points” in a multidimensional space. At the center of face space lies an average face, representing the central tendency of all the faces experienced. Typical faces, which are harder to remember, are located close to this average, while more distinctive faces, which tend to stick in our memory, are located further away (see Figure 2). Recent evidence suggests that the average face functions as a perceptual norm for coding
Face Perception

identity, highlighting what is distinctive about each face. The norm appears to be continuously updated by experience. A curious but telling side effect of this updating is that viewing a face biases us to see the “opposite” face, the so-called identity aftereffect (see Figure 2). Norm-based coding may be implemented by pairs of neural populations, one tuned to above-average and one to below-average values, for each dimension of face space. Common or typical faces produce little neural response with large responses reserved for novel faces, resulting in an efficient form of coding that may be used more generally in the visual system.

Development

Newborn infants can detect and visually orient to face-like patterns, and within days of birth they can discriminate faces, preferring their mother’s face to a stranger’s, attractive to unattractive faces, and faces looking at them rather than away from them. These observations suggest that newborns have some innate face-processing mechanisms, perhaps an innate face template or prototype combined with mechanisms that detect and orient visual attention toward stimuli that fit this template. Such mechanisms could direct face-like inputs to areas in the brain that become

Figure 1  Face recognition test


Note: It is easy to detect that the two upside-down faces in Row A differ, but harder to see that the two in Row B differ. Turning the faces up the right way makes the task easier, particularly for the pair in Row B, which differ only in how close the eyes are to the nose. It is likely that our sensitivity to the spacing of facial features is tuned to work best for upright faces because most of our experience with faces is in orientations close to upright.

Figure 2  Sample face space showing two identities, Dan and Jim

Note: Dan is a distinctive face that lies further from the average than Jim who is more typical. Also shown is an Average face and Anti-Dan, a face lying opposite Dan. Faces appear to be coded in relation to the Average face, which is continually updated by experience, resulting in face aftereffects. For example, staring at Anti-Dan for a few seconds biases participants to identify the Average face as Dan, much like staring at the downward motion of a waterfall induces the subsequent illusion that the stationary rocks on the bank are moving upward.
specialized for face processing (see below). The question of whether such areas have innate specialization for faces remains controversial.

During the first year of life, face-processing mechanisms appear to be “tuned” by exposure to faces. This results in perceptual narrowing analogous to the loss of sensitivity to phonemic differences that are absent from the infant’s language environment. In the case of faces, 6-month-olds can discriminate familiar from unfamiliar monkey faces as well as human faces, but by 9 months they can no longer discriminate monkey faces. A similar loss of sensitivity occurs for faces of unfamiliar races. In both cases, exposure to other-species or other-race faces prevents the loss. Disruption of early visual experience by cataracts produces deficits in holistic face processing and sensitivity to feature spacing that persist into adulthood, even when the cataracts are removed before 6 months of age. These deficits suggest that there are sensitive periods during which exposure to faces is essential for normal development.

Face recognition performance continues to improve throughout childhood. There has been considerable debate about whether this improvement reflects development of face-selective coding mechanisms or general cognitive mechanisms (memory, executive function, attention). However, recent evidence linking improved face recognition during childhood with expansion of face-selective cortex (see below) supports a face-selective component.

Reading Other Information From Faces

Faces are assessed on a variety of dimensions that modulate our social interactions. A salient one is attractiveness, which is assessed rapidly and with powerful consequences. Attractive faces activate reward centers in the brain and motivate sexual behavior and same-sex alliances. They also elicit positive attributions and treatment in a variety of settings (the attractiveness halo effect).

Contrary to a long-held belief that beauty is in the eye of the beholder and reflects arbitrary cultural conventions, recent evidence suggests that there is a biological basis for face preferences. People in different cultures show substantial agreement on attractiveness, and preferences emerge early in development, before cultural standards of beauty operate. Attractive facial traits include averageness, symmetry, sexual dimorphism (masculinity in male faces, femininity in female faces), a pleasant expression, good grooming, and youthfulness. Traits such as averageness, symmetry, and sexual dimorphism have been linked to disease resistance and developmental stability, and preferences for these traits may have evolved by sexual selection. In this case, the reproductive success of individuals who choose partners with these traits could be enhanced by direct benefits, such as reduced risk of contagion and better parental care from a healthier partner, and/or indirect genetic benefits, such as heritable resistance to disease in offspring. However, preferences for symmetry and averageness extend beyond faces and potential mates, suggesting a contribution of general information-processing mechanisms to such preferences.

People readily assess personal traits, including intelligence, health, competence, dominance, and trustworthiness based on facial appearance. Although not always accurate, these attributions can be useful predictors of behavioral outcomes. Indeed, one recent study found a substantial correlation between attributions of power-related traits in the faces of company CEOs and company profits.

Faces also provide cues to underlying emotional states, such as happiness, sadness, anger, fear, disgust, and surprise. These basic expressions can be recognized across cultures, although some have argued that only broader dimensions, such as valence (pleasure-displeasure) and activation (sleepy-hyperactivated), are universally perceived. Moreover, performance is often far from perfect, and emotion recognition is generally more accurate for faces of the same race and cultural group as the perceiver, similar to the own-race advantage typically found for face recognition. Facial expressions not only reflect genuine emotional experience. Expressions may be posed to provide social cues (e.g., polite smiles) or to intentionally deceive (e.g., masking a negative emotion with a smile). Our ability to detect deception from the face alone may be more limited than we generally realize, with performance often only slightly above chance.

Neural Mechanisms

The discovery of so-called face cells, which respond primarily to faces, in the inferotemporal cortex of monkeys, was an important milestone in the study of specialized neural mechanisms. More recent functional neuroimaging studies of the human brain...
suggest that face-selective tissue is distributed across occipitotemporal cortex, forming a core face-processing network (Figure 3). The occipital face area (OFA) in lateral occipital cortex (inferior occipital gyri in Figure 3) may form an initial, feature-based, face representation, which is projected to the fusiform face area (FFA) and superior temporal sulcus (STS). The FFA may individuate faces, whereas the STS codes changeable aspects of faces, such as emotional expression and gaze direction. The proposed network is consistent with the effects of brain damage, with damage to distinct regions producing deficits in different aspects of face processing (see below).

The core network interacts with an extended network of other cortical and subcortical areas to interpret the social cues available in faces (Figure 3). Subcortical structures are particularly important for rapid (less than 100 milliseconds), and probably automatic, perception of emotional expressions, as well as rapid detection of faces.

### Impaired Face Perception

Damage to the core face perception network can result in acquired *prosopagnosia*, where previously familiar faces, even one’s own, can no longer be recognized. This deficit can be remarkably selective for faces, since recognition may be unimpaired for nonface objects (e.g., birds) that, like faces, share a common structure and with which the person has considerable perceptual expertise (e.g., for a bird watcher). In contrast, object agnosia can occur in the absence of face-recognition deficits. This double dissociation between prosopagnosia and object agnosia suggests that distinct mechanisms may underlie face and object processing. Double dissociations have also been observed between deficits in recognizing identity and expressions, implicating distinct mechanisms for identity and expression processing (in FFA and STS, respectively), although these may operate on the same initial visual representation (perhaps computed in OFA). There may also be distinct neural substrates for the perception of different emotional expressions, with selective deficits in fear, disgust, and anger observed after damage to the amygdala, insula, and basal ganglia, respectively. Damage to these areas generally affects the experience, as well as the perception, of emotions, and this dual role may underlie our spontaneous mimicry of expressions and capacity for emotional empathy. Abnormal amygdala function is also seen in autism, another disorder associated with face processing and social cognition difficulties.

Deficits in face processing can also occur without apparent brain injury, in developmental or...
congenital prosopagnosia. It is not yet clear whether this constitutes a single disorder, because there is considerable variation in deficits. Nor is it clear whether it reflects an underlying pathology or simply the end of a broad distribution of face processing ability. Developmental prosopagnosia can run in families, exhibiting a simple pattern of inheritance consistent with a single dominant autosomal gene. Sex-linked genes are also implicated in human face perception, with deficits in face processing and social cognition occurring in the chromosomal disorder Turner syndrome (X-monosomy).

A Caveat

Although we can recognize many thousands of familiar individuals, sometimes even decades after seeing them, recognition can be poor for faces we have seen only once or a few times, with serious repercussions in forensic settings. For example, Jean Charles de Menezes tragically lost his life when he was mistaken for terrorist suspect Hussain Osman and was shot dead by Scotland Yard officers on the London Underground in 2005.

Gillian Rhodes and Linda Jeffery

See also Emotional Recognition, Neuropsychology of; Face Recognition in Humans and Computers; Facial Expressions, Computational Perspectives; Facial Expressions, Emotional

Further Readings


Face Recognition in Humans and Computers

The development of computer-based face recognition algorithms has intersected intermittently and productively with the study of human face recognition. Algorithms have offered insight into theories of human face recognition, and findings about the characteristics of human performance have filtered back into the development of algorithms. A primary point of intersection between human and machine studies has involved the question of how to represent and quantify the information in human faces. Evaluations of the strengths and weaknesses of both human and machine recognition have contributed to our understanding of the computational challenges involved in face recognition.

Representing Faces

Image-Based Representations

Face recognition for both humans and machines begins with an image projected onto the retina or sampled by a camera. Early approaches to machine recognition operated by extracting and quantifying the discrete features of faces (e.g., eyes, nose). A fundamental shift in this strategy occurred in the early 1990s with the use of principal component analysis (PCA) applied to face images. PCA is a statistical analysis that derives feature vectors from a set of input stimuli—in this case, a set of faces. The first measures used were simply pixel values extracted from face images. Applied to a population of faces, PCA creates a representation of individual faces in terms of a set of global features derived from the statistics of the face set. These global features consist of images that can be combined linearly to construct individual faces. From a psychological perspective, the connection of the feature set to the face population analyzed defines the “experience” of the computational model. The global features act as face descriptors that can specify categorical (e.g., sex, race) and identity information about a face. The computational components of PCA, therefore, implement a psychologically grounded face...
space model of recognition. By this account, faces are represented as points in a space, with the axes of the space defining features. The computationally implemented face space accounts for important findings in human recognition. The other-race effect, for example, is the well-known advantage people have in recognizing faces of their own race over faces of other races. This occurs because the space optimally represents faces similar to those used to derive the feature set (i.e., own race faces), consequently constraining the representation of other-race faces to be less than optimal.

Morphable Models

An important computational innovation came from changing the PCA input from images to image-based representations that are aligned with an average or prototype face. This representation supports morphing between individual faces and is therefore able to create faces that have particular properties useful for experimental manipulations. The alignment of faces is carried out using a set of landmark feature points (e.g., corners of the eye, tip of the nose). The shape of the face is coded in terms of the deformation of its landmark features from the average face. The reflectance or pigmentation information is analyzed separately in an aligned (shape-free) image space. A PCA is applied independently to the shape and reflectance information from a set of faces. This representation supports morphing between faces both in the reflectance and shape spaces allowing for the creation of synthetic morphed faces at arbitrary points in the space. The approach has been extended to operate on laser scans of faces that measure the three-dimensional shape of a head directly along with its reflectance map.

Psychological Results

Face representations based on deformations from an average face predict recently discovered high-level adaptation effects in human face perception. Face adaptation occurs after viewing a distorted face (e.g., expanded horizontally) for a few minutes, or less. The result is a bias to perceive a subsequently presented normal face as distorted in the opposite direction (e.g., contracted horizontally). Opponent adaptation effects have been found also for the sex of a face (i.e., adaptation to a male face biases the perception of an androgynous face to appear female), and for expression (i.e., adaptation to a sad face makes a neutral face appear happy). Moreover, these effects occur for face opposites with exposure to a synthetically created opposite face, a so-called anti-face, to facilitate the perception of the original face (see Figure 1). Opponent adaptation effects are consistent with face representations that directly code the deformation of an individual face from the average face.

Humans Versus Machines

Periodic international competitions for state-of-the-art face recognition algorithms organized by the U.S. government demonstrate an order of magnitude performance improvement for algorithms between the face recognition vendor test (FRVT) in 2002 and the FRVT in 2006. These algorithm evaluations require face identity matches for hundreds of millions of image pairs. To date, although the performance of algorithms across changes in viewing parameters (e.g., illumination) has improved, three-dimensional viewpoint changes are currently beyond the capabilities of most commercial face recognition algorithms.

Quantitative Comparisons

The level of accuracy achieved by humans recognizing faces is considered the gold standard to which algorithms should aspire. A recent comparison between humans and the algorithms from the FRVT-2006, however, showed that the best algorithms surpassed human accuracy at identity matching over changes in illumination conditions.

Figure 1  Example of a face (left) and its synthetically created opposite (i.e., anti-face)

Note: Exposure to the face facilitates the perception of the anti-face and suggests that human face representations reference an average or prototype face.
One qualifying factor in interpreting this result is that humans matched identity in unfamiliar faces. Human face recognition skills reach their highest level of expertise for familiar faces, showing robust recognition over changes in viewing parameters (e.g., illumination, viewpoint, and spatial resolution).

Qualitative Comparisons

Comparisons between humans and algorithms give insight into qualitative differences in the way humans and machines perform the task. In the early days of face recognition algorithms, it was possible to look directly at algorithms as models of human performance. This is no longer generally possible because the algorithms have achieved performance levels that make them commercially viable. As such, international competitions can no longer access the proprietary source code (i.e., algorithm) that achieves this performance. An alternative way to compare human and machine performance qualitatively is to analyze the pattern of errors they make. One way to test this is by fusing recognition estimates from several algorithms with estimates produced by humans. The rationale for this strategy is that fusion can improve performance only if there are qualitative differences in recognition strategies. Indeed, human-machine fusions improve performance substantially over the best algorithm operating alone. Empirically tested human-machine combinations can therefore exploit the strengths of both “systems” to perform more accurately.

Alice J. O’Toole

See also Face Perception; Facial Expressions, Computational Perspectives

Further Readings


Facial Expressions, Computational Perspectives

This entry will discuss specific mechanisms for decoding facial expressions, as well as different attempts to understand them with computational models. Decoding the intentions and the emotional conditions of others is an ability of vital importance for highly social species, including primates. Facial expressions are primary signals for interpreting emotional states; therefore substantial neural resources are dedicated to the perception and interpretation of facial features conveying meaningful expressions of emotion. It remains unclear, however, whether such resources have evolved specific computational mechanisms or have simply adapted generic cortical processing.

Facial Expressions: Background

With the publication of The Expression of Emotions in Man and Animals in 1872, Charles Darwin initiated the systematic, scientific study of facial expressions of emotion—on the encoding side. His work indicated that some emotions might have a universal facial expression and that animals, as well as humans, have a complex repertoire of emotional expressions. To understand decoding mechanisms, one should first consider to what extent expressions are a reliable, automatic manifestation of the internal state of the individual, which is still a debated issue. Facial expressions are often regarded as a veridical reflection of the signaler’s emotional state, an approach followed, among others, by Paul Ekman. In his work, Ekman has identified a set of basic expressions that he considers to be universal across cultures and to reflect physiological changes in the signaler, identifiable with emotions. If this position is accepted, then relatively fixed, perhaps even genetically specified, decoding mechanisms would largely suffice. An alternative view includes facial expressions in the wider discussion about the honesty of signals in animal communication. Within this approach, expressions are regarded as strongly modulated by the social context in which they are produced, and they function to manipulate the behavioral responses and the emotional state of the receiver. Therefore, they do not necessarily transmit honest information: Sometimes they can be
Facial Expressions, Computational Perspectives

largely veridical, while at other times they can be quite deceptive. Much more flexible decoding mechanisms are then required, to be honed by individual experience through learning and memory. Whatever system is used to interpret fine details in the visual input, based on a meaning-driven representation of repertoires of facial expressions, this system has to be continuously updated.

Are Facial Emotions and Facial Identity Processed Separately?

The functional model proposed by Vicki Bruce and Andy Young in 1986, one of the most popular “box models” from that period, had a strong influence on the study of face recognition, in particular with respect to the localization of perceptual mechanisms in the brain. The authors put forward the notion that facial expressions and facial identity are processed along distinct pathways, which diverge early on in the visual stream. The system devoted to expression analysis receives inputs from an early stage where the fine structural properties of the visual stimulus are already extracted, whereas a separate route is followed to abstract an expression-invariant description of the face, which will lead to the attribution of identity and other semantic information. This seemingly reasonable functional distinction between processing facial identity and emotion has since been assigned an anatomical basis, with the identification of cortical regions that are differentially involved in the visual analysis of distinct aspects of faces. Static (or perhaps virtually invariant) features of faces, including identity, seem to be especially encoded by the lateral fusiform gyrus, whereas the superior temporal sulcus seems to be involved in the representation of dynamic, changeable properties of faces, which include emotions, eye gaze, and lip movements.

Insights From Image Analysis Techniques

The complete divergence of these two systems at an early stage of the visual stream has been called into question by the observation that emotion and identity might partially share their neural representations. This more fully articulated notion is also supported by image-analysis approaches, based on the idea that techniques, such as principal component analysis (PCA) may simulate neural mechanisms of face perception and categorization. When such analysis was performed on a sample of faces with different identities and expressions, different sets of principal components were generated, some coding for identity, some for expression, and some for both. In this account, it does not appear necessary to resort to two anatomically separate systems to explain the independent perception of identity and expression demonstrated in previous studies, because both sets of properties can be extracted by a single, multidimensional system. The dissociation between them, in fact, may result merely from their distinct statistical properties, across an ecological range of face stimuli.

Apart from PCA, another image-analysis approach to facial expression recognition uses Gabor filters, with different frequencies and orientations. These units, which are assumed to model neurons with response properties similar to complex cells in the visual cortex, perform nonlinear edge detection at different scales and orientations. Methods have also been proposed that include in different layers Gabor filters, PCA, and categorization, to yield a more complex and hence presumably more biologically plausible model of facial expression recognition.

A Contribution From Attractor Dynamics

Whatever conclusion is reached in the localization debate, localizing processes in the brain does not, per se, clarify their nature. Similarly, describing the analysis of afferent (incoming) visual signals in terms of either PCA or Gabor filters tells us little about any decoding mechanism based on information that has been deposited in the cortex by previous experience. Such information is likely stored in interconnected neurons in a given cortical area (recurrent connections), or in top-down connections from more advanced to earlier cortical areas, rather than in the feed-forward connections that filter afferent signals.

Within a cortical area, the general notion is that clusters of statistically frequent or otherwise salient stimulus configurations produce corresponding attractor states—that is, preferred activity states toward which the network evolves. Such attractor states would be “carved” in the dynamics of local recurrent networks by Hebbian (i.e., associative) synaptic plasticity, in an unsupervised, effectively self-organizing learning process. This notion, at variance with the dominant connectionist paradigm of supervised (“back-propagation”) learning, has been investigated extensively with the aid of more brain-based network models, but only recently has it
received some empirical support in the observation of convergent single-unit activity. After passively representing a graded afferent signal for a few tens of milliseconds, single neurons are seen to converge in these experiments toward a more categorical response, putatively reflecting collective attractor states—states that attract the network to a unit of meaning. In the particular case of facial expressions, their graded intensity may often change over time, leading the network to slide along continuous attractors, using the same neural mechanism that has been proposed to underlie spatial navigation. If the network slides faster than the sensory signal changes, it will anticipate the forthcoming expression (or the forthcoming location, in navigation). Such a capability for prediction likely has high survival value. The superior temporal sulcus, implicated in expression decoding, is also known to be implicated in biological motion perception, and it has been interpreted as a generic generator of real-time predictions of forthcoming stimuli. In this account, expression perception might engage common cortical anticipatory processes, which predict the direction of the changes.

**Beyond Cortical Decoding**

Cortical visual mechanisms, however, might not be the only ones responsible for the identification of facial expressions. Subcortical visual pathways are believed to carry out a faster and coarser processing of faces, perhaps contributing to establish their cortical representation or perhaps largely bypassing the cortex to reach, for example, the amygdala.

More broadly, theories of embodied cognition propose that recognition of an emotional stimulus engages mental processes highly overlapping with those involved in the experience of the emotion. Brain regions that are commonly activated during an emotional state, including the amygdala, somatosensory, and motor areas, might then participate in the recognition of expressions of emotions. Impairments in the production of an emotion can in fact co-occur with deficits in expression recognition, and experimental disruption of neural activity in somatosensory areas impairs discrimination of facial expressions. It is doubtful, therefore, that we will be able to understand the decoding of facial expressions as if it were an isolated module of the mind.

Valentina Daelli and Alessandro Treves

**Further Readings**


**FACIAL EXPRESSIONS, EMOTIONAL**

The term *facial expression* refers to contraction of the facial muscles to produce an observable change compared to the appearance of a person’s face when muscles are relaxed. Emotional facial expressions are the facial muscle contractions occurring in the context of an emotional experience. Paradoxically, emotional facial expressions are referred to as *expressions* but they are far from a reliable readout of a person’s internal subjective state. A person may experience strong emotion without displaying any facial movement. Conversely, a person may display an emotional expression without feeling a corresponding emotion.

See also Computational Models of Emotion; Face Perception; Face Recognition in Humans and Computers; Facial Expressions, Emotional
In most facial expression studies, a substantial proportion of people display little facial activity, even under extreme circumstances. People frown when angry but also when concentrating on a difficult task; thus some facial movements do not have unique or invariant meanings. To further complicate matters, emotional facial expressions are not strongly correlated with physiological measures of emotional experience or with eliciting events and contexts. Nevertheless, emotional facial expressions are treated as meaningful by observers, and they clearly regulate interpersonal relationships, making their study important to social psychologists.

Inability to interpret the emotional meaning of certain basic emotional facial expressions is an important symptom of social dysfunction or cognitive impairment, an indicator of degenerative diseases, such as Parkinson’s or Huntington’s disease. Recognition tests have been incorporated into measures of social competence, emotional intelligence, and neurological functioning. Dysfunctional decoding of emotional expressions is also a symptom of disorders such as autism and schizophrenia. Impairment of expression and interpretation of expressions go hand in hand. Selective impairment of functioning suggests that the recognition of emotional expression is localized in the brain independent of recognition of facial identity.

This entry reviews the historical importance of emotional expression, current theories, and the scientific study of emotional expressions.

History

Early physiognomists studied the face in order to relate facial characteristics to aspects of personality. In the late 1800s, several neurologists, including Pierre Gratiolet, Sir Charles Bell, and Guillaume-Benjamin Duchenne de Boulogne, analyzed musculature and traced the anatomy of facial expression, describing the appearance of the face during certain emotional states. Charles Darwin relied on their work in his volume, *The Expression of the Emotions in Man and Animals*. Darwin’s work is considered the earliest systematic exploration of emotional facial expression. Darwin considered emotional facial expressions to be innate, arising from evolutionary processes. He suggested that there is survival value in the ability to signal important states to others, thus the ability to display and interpret expressions may have coevolved. He stated that expressions arise from associated habit and coordinated physiological arousal and proposed a theory of antithesis in which certain involuntary actions exist because they are the distinct opposite of other actions with survival value. Certain facial expressions may be lingering vestiges of behaviors that previously held survival value, such as eye widening during surprise or fear, originally to enhance vision but evolving into a signal to others.

Published in 1872, Darwin’s volume had little impact for over 100 years. Despite their reliance on the universality of smiling, frowning, and other expressive behavior when meeting new people in unfamiliar cultures, anthropologists such as Ray Birdwhistell and Margaret Mead insisted that facial expressions were socially learned behavior whose meaning varied from culture to culture. This view prevailed until psychologists Paul Ekman and Carroll Izard, encouraged by Silvan Tomkins, conducted separate studies demonstrating that certain emotional facial expressions were recognized consistently across widely varying cultures, including remote people in Papua New Guinea, isolated from Western influence. To date, hundreds of studies have replicated these first “universality” studies, finding highly consistent cross-cultural recognition of a set of basic expressions portraying surprise, joy, anger, sadness, and disgust. Expressions of fear and contempt produced less consistent results. In the 1990s, methodological challenges were raised by James Russell, but replications addressing Russell’s criticisms supported Ekman’s claims, and most theorists have now accepted the universality of the basic expressions.

In the 1970s, recognizing that systematic study would be difficult without a uniform way to describe the properties of facial expressions, Paul Ekman and Wallace Friesen developed the facial action coding system (FACS). Using FACS, observers assign numeric codes to distinct, observable combinations of muscle contractions, called action units (AUs). By systematizing description of the face, researchers have better controlled stimuli and better measured facial activity, resulting in a rich literature on facial behavior. While FACS was not the first or only coding system, it had the advantage of coding observed behavior rather than inferred emotional states and is now widely used to study a variety of facial activity. Similar coding systems for infant and primate facial expression have been developed. FACS has been
Facial Expressions, Emotional

implemented by computer recognition software, as have other systems for coding emotion directly in facial expression.

Current Theories of Emotional Expression

Emotional facial expressions are important to emotion research because they provide a measurable indicator of a subjective experience that is not directly observable. Studies of the nature of emotional facial expression have thus far been a battleground for those espousing differing theories of emotion. Unfortunately, most empirical studies of emotional facial expressions have difficulty distinguishing between competing theories, thus the field of emotion research remains contentious. Current theories of emotional expression may be categorized as (a) cultural relativist and constructivist theories; (b) universalist or basic emotions theories; and (c) perceptual, linguistic, and communication theories. Cultural relativist theories propose that emotional experience is socially constructed, thus facial expression can be expected to vary across cultures. Such theorists assume that facial expressions, or perhaps component movements of such expressions, become associated with emotional contexts through social learning. They suggest that the basic emotional expressions may represent cultural prototypes for the expression of emotion.

The universalist or basic emotions theories relate emotional facial expressions to a coordinated physiological response specific to an emotional state hard-wired by evolution into the brain. As neuroscientists have identified underlying emotion-response systems in the brain and body, basic emotions theorists have attempted to relate basic emotional facial expressions to them. Theories have changed with new findings in affective neuroscience. For example, with the advent of functional magnetic resonance imaging (fMRI), researchers attempted to isolate responses to emotional facial expressions to activation of specific areas of the amygdala. Empirical support for such approaches consists of cross-cultural studies of emotion recognition, studies showing that infants show basic expressions (before much learning can have occurred), that primates show basic emotional expressions, and that impairment of emotional response in those with neurological disorders is accompanied by impaired ability to produce and recognize basic facial expressions.

Communication theories, linguistic, and perceptual approaches apply ideas and terminology from cognitive science, psychophysics, and information processing to facial behavior. The encoding (creation) of an expression by a sender is viewed independently from its decoding by an observer (receiver), and the information content is analyzed. The characteristics of facial expressions may be analyzed in terms of pattern recognition, confusability, and signal detection. Researchers examine facial behavior in perceptual terms to identify cognitive processes used to recognize facial expressions. The search for salient cues signaling emotional meaning produced a debate about whether facial expressions are viewed holistically or as separable components. Evidence supports holistic processing, but the desire to link specific components of emotional expression to components of emotional experience persists in a variety of theories. Linguistic approaches suggest existence of facial expression “dialects” across cultures or try to link specific components of emotional expressions to linguistic emotion primitives, such as those suggested by Anna Wierzbicka. Component theories try to link specific facial movements with elements of emotional experience or the context in which it occurs.

Because of the weak association between facial activity and subjective states, some theorists have proposed a complete dissociation between internal feeling states and the instrumental use of facial expression to modify the behavior of others. Other theorists, such as Nico Frijda or Alan Fridlund, suggest that facial expressions signal behavioral intentions or action tendencies to observers but are not necessarily expressive of feelings.

Control of Emotional Expression

If facial expressions were to truly provide a readout of a person’s internal state, the privacy of one’s thoughts and feelings would be impossible to maintain. As it happens, smooth social interaction relies on the ability to simulate appropriate affect and to conceal inappropriate responses. The ability to manage emotional expressions arises from the existence of two neural pathways of innervation of the facial musculature. Neurons activated by the subcortical emotion centers of the brain produce spontaneous, involuntary emotional expressions. Neurons activated by the prefrontal regions of the brain and the
Facial Expressions, Emotional

motor cortex produce voluntary, deliberate facial expressions. These dual pathways produce a complex array of facial behavior. For example, a spontaneous smile might be dampened or controlled via pressing the lips together. A voluntary smile may be superimposed on muscle movement characteristic of anger. Although the face is the channel of nonverbal behavior most readily managed, people are rarely aware of what is happening on their faces from moment to moment. Deception can sometimes be identified via “leakage” of involuntary or spontaneous muscle movements evoked by emotion. In contrast to the dynamics of voluntary expressions, dynamics of spontaneous expressions tend to be slightly different (more symmetrical, of shorter duration), making it possible to detect deception if one is looking for it. However, there is little evidence that people use such cues in daily interaction, suggesting that most facial deception goes unnoticed. Paul Ekman has demonstrated that people can be trained to recognize subtle facial movements (microexpressions). Robert Rosenthal has found that commenting upon such nonverbal deception during social interaction is generally considered invasive and thus rude.

Emotional Contagion and Facial Feedback

During social interactions, an observer is likely to unconsciously mimic the facial activity of a companion. Studies suggest that when the facial muscles assume the position that occurs during an emotional expression, an individual will experience the corresponding emotional state. For example, if someone sees another person laughing, he will laugh too and feel enjoyment as a result of the laughter, even when there is no obvious reason for the laughter. This phenomenon has been long recognized and was explained by Carroll Izard as the result of neural facial feedback. Robert Zajonc suggested that muscle movement might cause alterations in blood flow to the brain resulting in the changes in mood experienced by those changing their facial expressions. More recently, the discovery of mirror neurons may provide an alternative explanation for the activation of emotional states via facial expressivity. Such neural mechanisms are now believed to be important to empathy and social regulation of affect. As with language, our interpretation of the emotional states of others may rely upon reexperiencing those states ourselves via motor mimicry. Infants have been found to mimic facial expressions from birth, and this may be important to early learning, both to control affect and to guide infant attention to the importance of the face during social interaction and communication. In adults, the mere presence of another person can calm a distressed individual, making social coping an effective method of regulating affect.

Sex Differences

Despite stereotypes, men and women are both emotionally expressive, although cultural norms may influence expression of certain emotions, such as fear for men or anger for women, especially cross-culturally. Voluntary smiling is required in certain occupations frequently held by women and has been found to occur more frequently among those with lower social status and power. In general, when experiencing the same emotional states, there are few differences in emotional facial expressions between the sexes beyond those attributable to socialization within a culture.

Nancy Alvarado

See also Emotion Recognition, Neuropsychology of; Facial Expressions, Computational Perspectives; Mirror Neurons

Further Readings

Flynn effect is the term commonly used to denote the observed increase in scores on IQ tests over time—for example, from one generation to the next. The Flynn effect can be understood as a birth-date effect in that average IQ rises with each increment in year of birth. The phenomenon is named for the New Zealand political scientist James R. Flynn, who described and quantified the rising scores in a series of publications beginning in the 1980s. The term Flynn effect was applied to the phenomenon by Richard J. Herrnstein and Charles Murray in their 1994 book, The Bell Curve. This entry begins by explaining the Flynn effect in terms of the conventions of IQ measurement. Then it describes the scope of the Flynn effect with respect to time and geography. It then specifies the evidence that led to discovery of the rising scores and describes some other data that have been used to characterize the Flynn effect. Then putative causes of the rising scores are considered along with theoretical and practical implications of the IQ increase. The entry ends with a brief note about the future course of IQ changes.

Psychometric Manifestations of the Flynn Effect

The scaling of commonly used IQ tests, such as the Wechsler tests and Stanford-Binet tests, ensures that the average score for a specified population is 100. If an IQ test uses a different scale, for example, a scale based on percentiles, the scores can be converted into equivalent IQs. In either case, the average IQ for a population would be set to 100. However, because older children tend to perform better than younger children and older adults sometimes perform worse than younger adults, it is necessary to calculate IQs separately for different age groups. Thus, even though actual test performance (raw scores) may differ substantially from one age group to another, the average IQ for each age group remains fixed at 100. Seen from this perspective the Flynn effect is an increase in raw scores from one cohort to the next. For example, today’s 30-year-olds are performing at a higher level than the 30-year-olds of a decade ago.

In the absence of any adjustment for the Flynn effect, IQ scores would continue to rise indefinitely. If the population average is to remain at 100, the normative data must be updated regularly. IQ norms accumulated today will be more stringent than they would have been a decade ago. For instance, achieving an IQ of 100 on a test normed in 2010 would require that a 30-year-old earn a higher raw score than would have been needed for a 30-year-old to achieve an IQ of 100 in the year 2000.

Magnitude and Ubiquity of the Effect

The Flynn effect is notable for several reasons. First, its magnitude is sufficiently large as to have significant impact wherever IQ tests are used. Depending on the test used to assess intelligence, average scores have been increasing at the rate of 3 to 6 IQ points per decade. Because of this increase in the population average, test performance that yielded an average IQ of 100 in 1960 would yield an IQ of only 70 to 85 if norms from the year 2010 were used. Consequently, an apparent decline in test performance from one age cohort to the next older cohort would be seen in the absence of any age-related deterioration of mental ability. Moreover, the Flynn effect is pervasive with respect to time, geography, and age group. On the basis of data from the United Kingdom, Flynn has inferred that the rise in scores can be traced back to the Industrial Revolution. Other data indicate that scores are rising on at least five continents. The Flynn effect can be observed in adults and children alike.

The Evidence of Rising Scores

An early indicator of rising IQ was found in data from the revised Wechsler Intelligence Scale for Children (WISC-R). During standardization of the revised test in 1971 through 1973, some children were administered the revised test along with the original WISC, which had been standardized about 25 years earlier. Scores were about 8 IQ points higher on the older test than on the revised test, which indicated that the new test had to be made more difficult (or the norms had to be made more stringent) to maintain a population mean IQ of 100. Flynn recognized this 8-point difference as clear evidence that the average IQ of American children had risen by 8 points in 25 years, or about 3 IQ points per decade.
Scores on the Stanford-Binet IQ test and various other mental ability tests have shown a comparable rate of increase over time, but scores on Raven’s Progressive Matrices (RPM) test have risen at twice that rate. The unadjusted score increases for Wechsler, Stanford-Binet, and Raven tests are illustrated in Figure 1. This differential rate of increase provides some insight into the mechanism or mechanisms underlying the Flynn effect. Scores on culture-reduced tests—that is, tests with relatively low demand for learned skills and knowledge—have increased much more rapidly than scores on tests that measure primarily the degree to which individuals have acquired word knowledge and other culture-specific information. Accordingly, scores from the nonverbal (performance) part of the Wechsler IQ tests have been increasing about twice as fast (4 IQ points per decade) as scores from the verbal part of the same tests (2 IQ points per decade).

There is no simple and universally accepted explanation for the rise in IQ. In a book about the Flynn effect edited by Ulric Neisser in 1998, several authors offered putative explanations that spanned a spectrum of possibilities ranging from improved nutrition to the influence of television and personal computers. Flynn has criticized all the explanations as inadequate to account for the persistence and pervasiveness of the rising scores. He has proposed instead that different factors have contributed to IQ increases at different times. Specifically, he has suggested that IQ increases in the United States prior to World War II can be attributed to the substantial increase in the quality and average duration of formal education that occurred in the country during the early 20th century. On the other hand, Flynn attributes more recent increases in test scores to an evolving emphasis on “on-the-spot” problem solving, as required in the contemporary workplace and in other realms of modern culture. According to Flynn, the importance of these dynamic and flexible mental skills (some of which are tapped by the Raven’s test) has overshadowed the value accorded traditional knowledge.

At this time, no one can specify the factors responsible for the Flynn effect, but Flynn and most
other commentators agree that the underlying causes must be environmental. The effect manifests itself as a difference between children and their parents and between children and their grandparents. Even if IQ were determined largely by genes, it is difficult to imagine a genetic mechanism that could produce a one-standard-deviation change in average IQ within three generations. This is perhaps the most intriguing and consequential aspect of the Flynn effect: An unknown set of environmental variables is producing a massive and ubiquitous rise in IQ.

Implications

The rapidly changing test scores necessitate caution in interpreting IQ test results if the norms are even a few years old. Moreover, the reach of the Flynn effect transcends IQ tests, as average performance on certain neuropsychological tests and memory tests also has increased rapidly. Age-group norms for adults on IQ tests and subtests, which are almost invariably cross-sectional, are confounded to varying degrees by the Flynn effect. Relatively low normative scores for older people may reflect largely their early birth dates rather than age-related loss of cognitive ability. Consequently, IQ tests have overstated the rate at which cognitive skill deteriorate during the life span.

Whereas some of these difficulties are of concern primarily to clinical psychologists, other ramifications are more general. For example, the provision of special education to a child may depend on whether the child’s IQ falls below a certain level, and the child’s IQ may depend on the recency of the norms for the IQ test. Now that the U.S. Supreme Court has prohibited the execution of convicts with intellectual disabilities, the Flynn effect has become critically important in certain capital murder cases. The fate of a convicted individual—execution or imprisonment—may hinge on whether his or her IQ falls within the range of intellectual disability. The obtained IQ, in turn, depends on the currency of the norms for that test. A current IQ test with up-to-date norms might cause the individual to be classified correctly as a person with intellectual disabilities. An older test with obsolete norms would yield an inflated IQ, and that inflated IQ might cause the convict to be classified as not mentally disabled and thus eligible to be executed.

In his 2007 book *What Is Intelligence?* Flynn examines theoretical dilemmas brought about by the rising test scores. The central theme is that intelligence cannot be what it traditionally is thought to be—namely, a unitary intellectual attribute that remains constant across the life span of an individual and across generations of a population. Instead, intelligence is a dynamic set of attributes that shift over time to reflect changes in the prevalent culture. Flynn points out that the Industrial Revolution set into motion a cascade of events that altered the world in many ways. Not only did more people receive more and better formal education, but thinking itself evolved from concrete, prescientific modes to the more abstract and conceptual modes that were stimulated by the growth of science and its spreading influence on the entire population. Brains did not become biologically better, but a rapidly changing culture demanded that each successive birth cohort meet a higher standard of abstract and conceptual thinking. Earlier cohorts were left behind the rising curve, and they still are.

Continuance Into the Future

Despite signs that the Flynn effect has diminished in Scandinavia, it is too early to predict a general flattening of the rising curve. If Flynn’s analysis of the effect is correct, mental abilities will continue to change over the long term to accommodate the changing demands of culture.

*Merrill Hiscock*

See also Heritability; Intelligence, Neural Basis; Intelligence and Working Memory

Further Readings


Folk Psychology

Folk psychology or commonsense psychology or theory of mind is the practice of ascribing psychological properties to people, such as belief, desires, emotions, and intentions. “Alice is angry,” “Juan is planning to stop Frank,” and “McCain thinks he is going to win the election,” are all psychological ascriptions. Folk psychology is “folk” because it is a practice that ordinary people engage in that does not presuppose any training in psychology or any other field of science. Psychological ascriptions are used to explain, predict, and understand others’ actions. It is common to think that the prototypical folk psychological explanation is a reason explanation. A reason explanation typically makes reference to a belief and a desire that together explain an action. For instance, “Frank thought Mary was drowning and he wanted to save her,” explains why Frank threw himself in the river. This entry concerns the main philosophical theories of folk psychology: theory theory and simulation theory, as well as some newer suggestions.

Both philosophers and psychologists have been interested in explaining what makes us capable of thinking of others and what they do in terms of such internal psychological states. Philosophers have mainly been concerned with explaining the adult capacity whereas most of the work in psychology on theory of mind is developmental. It is compelling to think that psychological states are given to us in experience. I have a concept of desire because I experience desire. The philosopher Wilfrid Sellars argued forcefully against this idea, maintaining that our concepts of psychological states are theoretical and form part of a theory of behavior. There is no sense in which “belief,” “desire,” and “intention” are given to us. It is against the background of such a criticism that current theories of folk psychology are best understood. Sellars’s view formed the basis of the theory theory of folk psychology, which, along with simulation theory, is among the most popular theories of folk psychology.

Theory Theory

Theory theory holds that we ascribe beliefs, desires, and feelings to people on the basis of a theory. This theory posits the existence of mental properties and consists in a great number of lawlike generalizations of the type “if a desires F and a believes that if she brings about G, she will have F, then a will attempt to bring about G, all other things being equal.” Proponents of this view, such as Jerry Fodor and Henry Wellman, maintain that it is because of such knowledge and its application that we are able to navigate the social world as successfully as we do. However, it is striking how few generalizations people are able to produce, even when given ample opportunity. This has suggested to some—for example, Shaun Nichols and Stephen Stich—that knowledge of folk psychology is best construed on the model of knowledge of grammar. Thus, folk psychological knowledge is tacit; it is innate in its elemental form, and explicit generalizations are likely post hoc systematizations of the underlying competence. Others, such as Heidi Maibom, have argued that knowledge of folk psychological theory is explicit. However, people do not express it in terms of generalizations, because their knowledge is of theoretical models, not generalizations.

Some philosophers argue that although we can formulate a theory to describe our psychological knowledge as expressed in our psychological attributions, we should not therefore suppose that ordinary people possess knowledge of such a theory. That is, people do not have an internal representation of this theory, although it appears to be implicit in their attributional practices. This position is less commonly held than the one described above.

Theory theory has come under attack, in part from simulation theory, which is discussed below, but also from newer, alternative theories of our ability to think about and forecast others’ behavior. Such newer approaches tend to be characterized either by a denial that psychological ascriptions, as practiced by ordinary people, are truly explanatory or predictive or by a rejection of the idea that theory theory explains our ability to engage in sophisticated social behavior.

Some theory theorists argue that folk psychological theory is radically false. For instance, Paul Churchland has argued that the theory represents a stagnant or degenerate research paradigm. It has remained essentially unchanged since the Greek epics, it fails to explain a range of psychological phenomena, including learning and sleep, and it is unlikely ever to integrate with, or be reducible to, more successful scientific theories. As such,
we should accept that people do not, in fact, have beliefs, desires, and so forth. These are merely useful fictions. This idea, known as eliminative materialism, is not generally accepted, although most philosophers accept that mental states are probably somewhat different from how folk psychology conceives of them.

Theory theory is closely related to the representational theory of mind—that is, the theory that mental states are representational states. In the developmental literature, people, like Alison Gopnik and Andrew Meltzoff, have argued that one can observe theory change in children—that is, the transition between a picture of the mind as passively reflecting reality to one in which the mind actively represents the world. This is particularly striking once children come to pass the false belief task, which tests a person’s ability to understand that someone else could have a false belief—a belief that does not correspond to reality. Some detractors argue that children have a representational theory of mind very early, and failure to succeed on the false belief task is a question of application or performance, not competence. Children may, for example, have problems inhibiting the default response to belief questions, which is to ascribe to the agent beliefs that correspond to reality as they see it.

Simulation Theory

Originally proposed as an alternative to theory theory, simulation theory maintains that we ascribe psychological properties to people by putting ourselves in their shoes, imagining what we would think and feel under those circumstances and then ascribing those thoughts and feelings to others. Proponents of this theory include Jane Heal, Robert Gordon, and Alvin Goldman. Simulation does not require us to have anything like knowledge of a theory, or at least not knowledge of a psychological theory. In the simplest case, we project ourselves wholesale into the position of the other and through this imaginative step come to have certain thoughts, desires, and emotions, all within the scope of the simulation. Once we have acquired the information that we are interested in, we end the simulation and take up our own perspective again. In more complex cases of simulation, we import into the simulation information about relevant differences between us and the target. We then imagine, within the context of the simulation, that we have certain beliefs, desires, ways of thinking about things, and so on. The final result of a simulation is usually thought to be a mental state attribution or an action prediction. Because you change your perspective at the end of a simulation, simulating others does not make you want to act like you imagine they would want to act; it merely gives you information about how they would act. There are, however, some suggestions in the literature that simulating others’ emotions leads to empathizing with them, which suggests that simulation often goes beyond giving you information about others.

Simulation theory has been criticized, among other things, for its reliance on pervasive similarities between simulators and targets. The trouble is, some say, that simulation would have us be wrong about others very often because others simply are not that similar to us. But there are reasons to think that we are wrong about others. According to William Ickes, we are only accurate in our psychological ascriptions to strangers 20% of the time. We also tend to overestimate the degree to which others are similar to us. Other criticisms focus on the apparent reliance on the introspective availability of psychological states, which someone like Sellars famously denied. But most modern theory theorists are likely to hold that we arrive at self-ascriptions via a different route than third person ascriptions anyway, and it is not so clear that this route would prove unavailable to simulation theorists.

Where traditionally theory theory and simulation theory were competing positions, most proponents are now willing to grant that both accounts are descriptively true; sometimes we simulate others to understand them better, and sometimes we use theory. There are a variety of different accounts about how the two differ, when the different approaches are used, and so on. Only some insist that simulation is absolutely essential to understanding others.

Alternative Theories of Folk Psychology

More recently, a number of philosophers have expressed discontent with the traditional conceptions of folk psychology just mentioned. Many theorists now doubt that our ability to think of others in terms of their thoughts, desires, feelings, and intentions is required for the variety of social interactions that we engage in on an everyday basis. One problem is that many animals that do not have such...
knowledge are perfectly capable of engaging in quite sophisticated interactions with conspecifics (i.e., others of the same species). Another is that it does not seem true to everyday experience. Our everyday predictions and explanations of others often do not seem to rely on ascriptions of psychological states at all. It is therefore now increasingly popular to claim, as José Bermúdez does, that psychological ascriptions are usually only pressed into use when people act in problematic or unexpected ways. The rest of the time, we make use of other structures. For instance, we may make use of our ability to see behavior as goal directed (perceptual animacy) without thereby having to ascribe desires to people. We may also use our knowledge of social systems to predict what someone is going to do in virtue of his embeddedness in a social structure, as Maibom has argued elsewhere. For instance, we think that someone is going to give us a menu, not by reflecting on any of his beliefs or desires, but by observing that the person is a waiter.

While some proponents of alternative theories are happy to keep room for traditional psychological ascriptions playing some role in explanation and prediction, people, like Kristin Andrews, claim that such ascriptions only help us understand others, but not because they provide accounts that bear any interesting resemblance to explanations or predictions as we find them in the sciences. By contrast, Adam Morton argues that if we were to base our decisions of what to do on predictions of others’ actions, it would lead to insoluble problems, since often what others are likely to do depends on what they think we are likely to do and so on. There are, therefore, many reasons to think that social interaction does not rely on psychological ascriptions—whether arrived at by theory, simulation, or something else—to the extent that people used to think. It should be noted, however, that some such alternative accounts take themselves to be modern forms of more traditional accounts—for example, theory theory accounts.

In conclusion, the landscape of folk psychological theorizing is changing. Primarily, people are proposing less intellectualized accounts. That is, rather than assuming that relatively advanced knowledge of psychological states lie behind our social prowess in general, the more popular view now is that what underwrites this prowess are often much simpler abilities, most or many of which do not deploy anything like representations of psychological structures.

Having said that, few accounts reject theory theory and simulation theory outright. The move is rather toward more eclectic approaches.

Heidi L. Maibom

See also Eliminative Materialism; Realism and Instrumentalism; Representational Theory of Mind; Social Cognition

Further Readings


**Freedom of Action**

It is commonly supposed that human beings have a unique ability to freely control their actions and decisions and can be held responsible for their actions on account of this capacity. But there has been much debate about what such freedom involves and whether it even exists. This entry will examine the nature of freedom and the potential threat posed to it by determinism and will go on to look at the two main opposing viewpoints within the debate: the view that freedom and determinism are compatible with
one another and the view that freedom and determinism are incompatible with one another.

The Debate About Freedom

The dispute arises when we consider what is meant by \textit{freedom} and what is required for humans to possess it.

The Concept of Freedom

When we use words such as \textit{freedom}, we seem to have a number of related concepts in mind. Consider the following three:

1. Free agents are generally taken to possess a capacity to choose between various alternatives, or to have futures that are \textit{open}.
2. Free agents are generally taken to have a particular sort of control or autonomy, which allows them to decide \textit{for} themselves: You might suppose the future is not merely open, but it's \textit{up to you} which actions you perform and hence which of those alternative futures really occur.
3. Freedom is often thought to ground our view that agents are the source of their own actions in a certain way: It allows us to attribute actions to agents in a way that is often said to ground an agent's responsibility for those actions.

Threats to Freedom

Various phenomena might be thought to threaten our freedom, but in recent years, the debate has focused almost solely on the potential threat from \textit{causal determinism}: This is the theory that prior causal conditions, along with the laws of nature, causally determine all future events; that the future is fixed by the past and the laws of nature.

If we had a complete statement of all the facts about the universe at some moment in the past and a complete statement of all the laws of nature, then the theory of causal determinism asserts that these statements together would entail every true statement about present and future facts.

The Threat of Determinism

While many people find it intuitive to suppose that if determinism is true, then human action cannot be free, any direct inference from determinism to a lack of freedom would be too quick. Some further assumptions are required to generate this threat. These assumptions tend to rest on claims about the relation between freedom and \textit{alternative possibilities}.

If determinism is really to threaten freedom, then we need to accept the following claims:

1. Freedom requires alternative possibilities.
2. Alternative possibilities are ruled out by determinism.

These assumptions may instead be stated in terms of an ability to do otherwise:

1. Freedom requires the ability to do otherwise.
2. The ability to do otherwise is ruled out by determinism.

Given the plausibility of these further claims, along with the plausibility of determinism and our natural conviction that our actions are free, the problem might be stated as a dilemma. We have the following, seemingly plausible claims, which cannot \textit{all} be true at once:

1. Alternative possibilities are ruled out by determinism.
2. Freedom requires alternative possibilities.
3. Determinism is true.

The various different stances within the free will debate can be seen as possible ways of resolving this dilemma.

One way to deal with the dilemma is to claim that determinism does not really pose any threat to freedom at all. This stance is known as \textit{compatibilism}. The compatibilist generally focuses on denying either the first or second of the claims above: those that are needed to generate the threat to freedom posed by determinism.

But some theorists accept that freedom is threatened by determinism and try to resolve this dilemma by denying the third or fourth claim. These theorists are known as \textit{incompatibilists}. Those who deny that determinism is true and maintain that we \textit{do} in fact have freedom are known as \textit{libertarians}. Those who instead deny that we have freedom are sometimes called \textit{pessimistic incompatibilists}.
To understand the debate about freedom, it will help to consider these basic stances on the problem in more detail.

**Compatibilism**

Compatibilists tend to deny either Claim 1 or Claim 2 above: They deny either that alternatives are ruled out by determinism or that they are necessary for freedom.

**Traditional Compatibilism**

Traditional compatibilists claim that agents are free so long as they can do what they want to, unencumbered by constraints and impediments. This is the sense in which it must be possible for agents to do something else: There must be no external constraints stopping them from performing such an action if they want to.

This condition could be met even if determinism is true, so this form of compatibilism denies the first of our claims: that determinism rules out alternative possibilities. Alternatives of the sort this view takes to be relevant are not ruled out.

Some compatibilists appeal to a *conditional analysis* in support of this. On such an analysis it is argued that the following claims should be considered equivalent:

a. The person could have done otherwise.

b. The person would have done otherwise, if he or she had wanted to.

Sometimes *wanted* is swapped for *chosen* or *intended* or *decided*. But the basic claim is that there is a sense of “able to do otherwise” where this is conditional on the wishes or the choices of the agent. And this sense of able to do otherwise is not ruled out by determinism.

Incompatibilists often express dissatisfaction with this analysis, arguing that agents will only *genuinely* be able to do otherwise if, further to the fact that they would have done otherwise if they wanted to, the following is also true:

c. They could have wanted to do otherwise.

If the agents could not possibly have had different desires or made different choices to begin with, it is asked how it could be true that the agents are able to do otherwise?

But this objection seems to invoke a broader sense of able to do otherwise, which is not the sort that compatibilists accept as relevant. Traditional compatibilists argue that the important point is about whether the agents can act on the basis of their own will, and not about whether their will could be different in the first place.

Whether this really *does* capture the ability we care about when we ask questions about an agent’s freedom, however, continues to be a moot point. Incompatibilists typically differ from traditional compatibilists in seeing the freedom of our actions as being explained, ultimately, by a freedom to *decide or will* differently and not by a capacity to merely act on the basis of our predetermined will. There is much dispute about this, and both compatibilists and incompatibilists claim that their own view captures our *ordinary* conception of freedom more accurately.

**Frankfurt-Style Examples**

A more recent trend within compatibilism has been to accept Claim 1 above and to deny Claim 2: to accept that determinism rules out alternative possibilities but deny that alternative possibilities matter for freedom. This is supported by a certain type of thought experiment first constructed by Harry Frankfurt and often called *Frankfurt-style examples*.

A typical Frankfurt-style example goes as follows: Black wants Jones to perform a certain action. He has an ingenious device that will manipulate Jones’s nervous system and brain processes, making him perform the action in question. But Black is excellent at judging people’s intentions, and he will only bother doing this if he judges that Jones is not going to perform the act of his own accord. However, Jones decides to perform the act of his own accord, so there’s no need for Black to intervene.

In this example it looks as if Jones is free—or at least responsible—despite the lack of alternatives. Hence alternative possibilities seem to be irrelevant.

In response to this, incompatibilists tend to question whether this example really does rule out alternative possibilities. If Black can judge what Jones is intending, and he chooses to manipulate him based on this, is it not still possible for Jones to intend something else in the first place? If so, then there is still an alternative possibility present, since Jones could form an alternative intention. The
incompatibilist can appeal to this alternative possibility in explaining our intuition that Jones is free.

Compatibilists tend to question whether such narrow alternatives—often called “flickers of freedom” could be enough to ground a significant ability to do otherwise: What if Black puts a device in Jones’s brain that responds to minute differences in earlier patterns of neuron firings and can predict (and prevent) his decision on this basis? Would flickers of this minute sort constitute a robust enough “ability” to do otherwise?

Many incompatibilists argue that they would. There is very little agreement about which kinds of alternative should “count” and why. So the success of Frankfurt-style examples remains a matter of dispute.

Incompatibilism

Incompatibilists accept both that freedom requires alternatives and that such alternatives are ruled out by determinism, and hence freedom and determinism are not compatible with one another. They are therefore committed to denying Claim 3, that determinism is true, or Claim 4, that we are free.

Libertarians

Libertarians argue that human actions are free and that determinism is false.

“Agent causation” accounts maintain that agents must be the cause of their own actions, as opposed to those actions being caused by prior events, some of which are external to the agent. But such accounts are often criticized for invoking a rather strange conception of causation: In nature generally, only events tend to cause other events. In this view, “substances” can cause events, and this seems difficult to square plausibly with an ordinary conception of causation.

Not all libertarians see freedom in terms of agent causation, however. Noncausal views deny that the ability to act freely is a capacity to “cause” one’s own actions at all, as opposed to a power of some other sort. Event-causal views deny that we have to give this sort of causal role to agents themselves rather than to prior events but nonetheless maintain that alternative possibilities are important for freedom.

Libertarians all seem to run into another problem: It’s sometimes questioned whether freedom can be reconciled with indeterminism any more than it can be reconciled with determinism. Perhaps human actions could not be free if they are determined by prior causes, but it’s not clear whether they could be free if they simply happen by chance either. This seems hard to reconcile with the sort of control we ordinarily suppose free agents to possess.

Pessimistic Incompatibilism

Other incompatibilists accept that determinism is true and that it rules out freedom. They instead reject Claim 4 and argue that human actions are not free after all.

But it’s common to suppose that such a view would require us to give up some very fundamental commitments. For instance, we would perhaps have to get used to a world in which no agent could ever be held responsible for his or her actions and in which praise and blame are considered irrational. It can be difficult to imagine how standard social and legal institutions would function in the absence of concepts such as praise and blame or punishment and reward.

Some philosophers argue that it makes no sense to try to “give up” concepts of this sort, because they are so deeply ingrained in our attitudes to one another: Arguably, even the most basic social interactions require that people see one another as responsible agents. However, others have questioned whether this is really the case and whether such a conclusion could justify our attributions of responsibility.

Nadine Elzein

See also Explanation of Action; Mental Causation; Phenomenology of Action; Philosophy of Action; Voluntary Action, Illusion of

Further Readings


FREGOLI DELUSION

The Fregoli delusion is a rare delusional belief that strangers or unfamiliar people are in fact persons familiar to the patient in disguise. Usually, the patient believes that a persecutor or an impostor has changed his or her appearance by taking the form of a stranger in order to deceive the patient.

This entry summarizes the main characteristics of the delusion, its history, its nosological setting (i.e., its description and classification), and the etiological or causal hypotheses addressing the delusion.

Overview

History

The delusion was first described by Courbon and Fail in 1927 and was named after Leopoldo Fregoli, an actor at the time who was famous for his disguises on stage. The authors described a 27-year-old woman who believed she was being persecuted by two actors whom she often saw at the theater. The woman believed the actors pursued her closely, taking the form of people she knew or met.

The Fregoli Delusion: Nosology

The delusion has been categorized as one of the four main and classical delusional misidentification syndromes, the other three being the following:

a. Capgras syndrome (in which the patients believe that a person, usually a relative, has been replaced by a double)
b. Intermetamorphosis (in which the patients believe that persons in their environment change places with one another)
c. The syndrome of subjective doubles (in which the patients believe that other persons have been transformed into themselves)

The four syndromes are grouped together since they often co-occur and interchange in the same patient as well as because they present the same basic themes—namely, the negation of true identity and the belief in doubles (sosies).

In the classic form of the Fregoli delusion, patients were thought to make a “hyperidentification.” Although they recognize the forms of the strangers, they identify those people as other people, specific persecutors. In contrast, patients with the Capgras delusion are thought to “hypoidentify” their relative by denying his or her identity. Delusional misidentifications have also been reported to concern misidentifications of objects or places and therefore are thought to be linked to another delusional belief sometimes found in neurological patients, reduplicative paramnesia (a delusional belief that a place or a location exists simultaneously in two different places or that it has been relocated to another site).

Etiological Hypotheses

Fregoli delusion as a symptom usually occurs in psychotic illnesses and especially in schizophrenic illness, schizophreniform psychosis, and schizoaffective and affective disorders, though it has been reported to occur as a manifestation of various organic brain disorders, such as head trauma, ischemic brain damage (i.e., from a stroke), Alzheimer’s disease, Parkinson’s disease, epilepsy, or metabolic diseases. Delusions of misidentification in general appear to be associated especially with organic lesions affecting mostly limbic structures and, to a lesser extent, both the frontal and parietal lobes. Contemporary accumulated evidence suggests that right hemisphere lesions predominate in the etiology of delusional misidentifications.

Fregoli delusion, for example, was reported in a patient following abruptly after ischemic brain damage that involved the right fusiform gyrus and face specific area; furthermore it is estimated that up to a third of all patients with Alzheimer’s disease display the phenomenon at some point during the course of their illness. Delusional misidentification syndromes therefore have been used as an exemplar pathological phenomenon for the study of delusion formation. However, despite the accumulated evidence for the importance of organic factors in the etiology of these syndromes, physiological data alone, according to present knowledge, cannot explain the “delusional” nature of the belief, and therefore, the pathogenesis of the delusion remains unclear.

Feelings of familiarity have been considered crucial to the pathogenesis of these specific delusions by many authors, and the relevance of affect in relation to cognitive functions is regarded as of major significance in explaining their nature.

Cognitive Neuropsychiatric Hypotheses

Delusions of misidentification are considered an ideal and promising ground for testing theories and models that attempt to explain the formation process
of delusions more generally. Cognitive neuropsychiatric hypotheses suggest that these specific beliefs could derive from defects at different stages of an information-processing model for face recognition, which was introduced to explain prosopagnosia (the inability to recognize faces of significant others). In contrast with the Capgras delusion, which is considered to occur when people lack affective confirmation of the person they see while receiving the appropriate semantic information, in the Fregoli delusion, hyperfamiliarity feelings or heightened affective responsiveness while seeing a stranger is considered to be of major importance.

The Model of Face Recognition

Cognitive neuropsychiatric explanations of the Fregoli delusion employ an information-processing model of face recognition that deconstructs the operation into three essential stages: (a) an initial structural encoding, (b) the excitation of units sensitive to the unique characteristics of each face, and finally (c) links to other multimodal nodes that access biographical and episodic information about people. The genesis of the Fregoli delusion has been hypothesized to lie in the last stage. At this stage, hyperactive person nodes are said to fire in the presence of other similar people.

The Two-Factor Account

The above-mentioned hypothesis stumbled on its inability to explain why the patient does not reject the idea of transformation or substitution as irrational or illogical, which would normally happen when a person without delusions misidentifies someone else. Since this model alone cannot explain the delusion, another factor, which is responsible for evaluating decisions, seems requisite. This second factor, the decision-making stage, is considered to be defective as well. The two-factor account has been used to explain a number of monothematic delusions, aside from the delusions of misidentification.

The Identification Disorder Hypothesis

Other hypotheses suggest that the main disturbance in the case of these delusions is not in the face recognition process but rather in a stage responsible for attributing a unique identity to surrounding persons and objects. This stage is considered to be indispensable for integrating past and recent information to allow identification to be constant over time. It prevents small differences in factors such as appearance, lighting conditions, or anomalous feelings of familiarity from interfering with proper identification. Delusions of misidentification are therefore regarded as an identification disorder. In the Fregoli delusion in particular, according to this hypothesis, misidentifications could be triggered by various situations independent of the persons around the patient, such as feelings of significance, altered affect, or associative memories. Fregoli patients may trust these misidentifications and not see an alternative since they appear to lack the healthy ground that serves the identification process and aims to counterbalance these experiences.

These explanatory hypotheses and models should be tested to advance our understanding of the formation of this fascinating delusion.

Management and Treatment

Management and treatment of the Fregoli delusion, as in the rest of the delusional misidentification syndromes, requires a thorough investigation aimed at identifying an organic lesion or a more widespread pathology. Treatment is mainly symptomatic and includes antipsychotic medication, as well as the treatments for the underlying condition, whether psychiatric or organic.

Maria M. Margariti

See also Capgras Delusion; Delusions; Face Perception; Face Recognition in Humans and Computers

Further Readings


FREQUENCY EFFECTS IN WORD RECOGNITION

A word's frequency is a measure of how commonly it is used. Most often, it is expressed as a relative frequency, such as occurrences per million words in a large corpus of written or spoken language. Generally, higher frequency words are recognized more quickly and/or accurately than lower frequency words, although processing is simultaneously affected by many lexical characteristics (such as word length, spelling-sound regularity, and number of phonologically and semantically similar words). While corpus frequency provides only an estimate of the average person's experience, it correlates strongly with an individual's subjective frequency estimates and is one of the best predictors of word recognition facility. This entry briefly reviews basic phenomena of frequency effects and mechanisms proposed to account for them.

Phenomena

In a basic study of frequency, the experimenter tests performance on sets of low- and high-frequency words. The experimenter must operationalize “low” (perhaps 1 to 10 occurrences per million words) and “high” (perhaps greater than 50 or 100 occurrences per million, but specific levels vary between studies). Log frequency predicts performance better than raw frequency, such that the effect of a constant difference in raw frequency diminishes as word frequency increases (e.g., a difference between 1 and 20 per million has a stronger effect than a difference between 101 and 120 per million). One can also study frequency using a regression approach and test performance on words with a continuous distribution of frequencies. In such studies, frequency tends to account for 5% to 15% of the variance in performance when other lexical characteristics are controlled.

The basic frequency effect is easily replicated, both in the rarefied conditions of isolated word processing as well as in more ecologically valid tasks, such as reading and following spoken instructions to interact with objects in a visual display. In reading, for example, fixation time on a word is inversely proportional to its frequency. Frequency can also influence performance on processing speech sounds. For example, categorical perception of a spoken phoneme continuum such as /b/ to /p/ shifts if the continuum is between a high- and a low-frequency word, as in “best-pest” (shift toward /b/) versus “pray-bray” (shift toward /p/). Frequency also interacts with other lexical characteristics such that effects of other variables (e.g., spelling-sound regularity and neighborhood) are more pronounced in low-frequency words than in high-frequency words.

Basis

A very basic question about the basis of frequency effects is whether the causal factor is cumulative frequency of exposure (as implied by using corpus estimates) or some aspect of experience. For example, high-frequency words also tend to be words that are acquired early by children, suggesting that early age of acquisition might afford a privileged status in memory and raising the possibility that frequency effects might best be understood as effects of age of acquisition. However, behavioral studies and simulations with connectionist models have established that age of acquisition and cumulative frequency both influence lexical representations. A related question concerns the impact of long-term cumulative frequency versus that of recent changes in frequency. Although changes in recent experience can make frequency effects disappear quickly (e.g., as low- and high-frequency words in a closed or fixed set are repeated, performance on low-frequency words quickly catches up to performance on high-frequency words), such effects dissipate without continued support for short-term changes in frequency. This suggests differential weighting of long-term and recent experience.

Three Classes of Proposed Mechanisms

One possibility is that frequency constantly and directly affects a word's availability. For example, the lexicon might be organized like a file drawer with entries sorted by frequency. Similar behavior would be predicted in a connectionist framework by making the “resting level” (default level of activation in the absence of bottom-up support) of the node or network representing a word proportional...
Frequency Effects in Word Recognition

First, a method for tracking the time course of spoken language processing was developed by Michael Tanenhaus and his colleagues. In their “visual world paradigm,” subjects’ eye movements are tracked as they follow spoken instructions to move objects in a visual display. Fixations are remarkably strongly time locked to fine-grained phonetic details, with only about a 200 milliseconds delay between a phonetic detail and its impact on gaze. When this technique was applied with low- and high-frequency words, the time course was like that predicted by a connection strength mechanism: Frequency effects were immediate but initially subtle and strengthened as a word was heard. Converging evidence has been reported for both visual and spoken word recognition in a dual-task paradigm where frequency effects are present prior to the occurrence of a “dual-task bottleneck” (the point where performance on each task begins to suffer from interference from the other). Performance prior to such a bottleneck is presumed to depend only on highly automatic processes, suggesting an early and automatic locus of frequency effects.

James S. Magnuson

See also Word Recognition, Auditory; Word Recognition, Visual

Further Readings


GENDER DIFFERENCES IN LANGUAGE AND LANGUAGE USE

This entry reviews the distinction between sex-exclusive and sex-preferential differences in language, then considers how preferential differences in language use acquire social meaning. Typical trends of gender differences in language variation and language change are outlined and possible motivations considered.

Sex-Exclusive Versus Sex-Preferential Differences

In the study of women’s and men’s ways of talking, it is common to distinguish between gender (which is considered a social construct) and sex (a biological fact); this entry will try to be consistent with this distinction. In practice, researchers often blur the two; for example, they generalize about gender differences based on samples that have been selected on the basis of sex differences.

There are very few sex-exclusive differences (i.e., cases where men and women use completely different forms or languages at all times) in language use anywhere in the world. Early European reports from the Caribbean about communities where the sexes spoke different languages reflected misunderstandings of the linguistic variation in those communities. Communities such as the longhouses of the Vaupes, where linguistic exogamy is practiced (marriage partners are chosen on the basis of speaking a different language from their partner’s community), are perhaps the best and only well-documented example of sex-exclusive linguistic practices. However, even there, women and men can understand and sometimes speak their partner’s primary language; the restrictions are social ones related to appropriateness of use.

The vast majority of gender differences in language use are not sex exclusive; instead, they are sex preferential. That means male and female speakers make more or less use of the same pool of linguistic resources. What is perceived as gender differences reflects our perception about the probability with which a given form is likely to be produced by a female or male speaker and/or social ideologies about whether a particular way of pronouncing a word, the expression of particular interpersonal stances through language, or the use of a particular speech act is normative for men or women. These generalizations speak of typical patterns across a wide range of men and women in a speech community; it is therefore unsurprising that there is intraindividual and interindividual variation. Speakers may use phrases or pronunciations that are normatively associated with their sex with different frequencies in different social contexts, and some speakers use linguistic forms and routines considered normative of their gender more often than others do.

Social Meaning of Gender Differences in Language Use

Research on gender and language has moved through several stages: from a focus on gender difference, to a focus on the politics of dominance, to
a focus on the social meaning of expressions of both difference and dominance. This reflects an increasing appreciation of the complexities of difference and power as social and interpersonal phenomena. For example, supposedly “men’s” and “women’s” forms in Japanese can in fact be used by either sex (in other words, so-called women’s language is a misnomer). A more economical and comprehensive analysis of these forms is in terms of the social stance the forms express. Because some forms are associated with stances of assertiveness and others with stances of softness, and because Japanese culture normatively associates softness with women and assertiveness with men, the use of the variants tends to be associated with speakers of that sex. However, men can use soft variants if the social context requires it, and women can use assertive variants for a particular social effect. Recent research in the framework of evolutionary psychology is a notable exception to the trend toward understanding social meaning; its focus on difference (even where gender differences are slight) and the deterministic framing of difference as evolutionary necessities have been criticized by sociolinguists. Deborah Cameron argues against accepting evolutionary accounts of differences in women’s and men’s language because (a) they often presuppose the object of enquiry—for example, that women talk more than men (notwithstanding plenty of evidence to the contrary)—and (b) they selectively frame such generalizations; for example, the proposal that men talk less because prehistoric hunting patterns required silence ignores the fact that hunting large game was probably a (rare) group activity requiring as much social organization as the more common (female and male) activities associated with gathering.

Common Patterns of Gender Difference in Variation and Change

Notwithstanding inter- and intraspeaker variability, we can draw some generalizations about gender differences in language. These are most useful when they illuminate core problems of linguistics, such as the relationship between social structure and language change. All languages are constantly changing. A well-attested pattern is for female speakers to lead men in ongoing changes; women use more of the innovative variants that are used most often by younger speakers. The much-quoted and highly misleading generalization that women are linguistically more conservative than men applies only to some, but not all, stable variables—that is, parts of the linguistic system where the variation is not involved in change. This is true for the stable alternation between an alveolar and a velar nasal in finding [fəndɪŋ] versus [fəndɪŋ]: The standard velar variant is generally produced more by women than by men; the vernacular alveolar variant is more likely to be produced by men. But even this generalization interacts intimately with social factors such as level of education. Work on stable variables in Arabic-speaking communities finds the reverse: Classical Arabic forms are more likely to be produced by men than women.

Motivations for Gender Differences

The reasons why women typically lead men in ongoing change are unclear. One proposal is that because women continue to be the primary caregivers for small children, children perceive what women do to be the norm and perpetuate this. However, this account leaves the fairly steady pace of language change across generations a mystery. It might explain why little boys end up using more “advanced” variants than their fathers, but it’s less clear how it explains why little girls also end up using more advanced variants than their mothers. Another proposal is that language serves as a more potent symbol of social capital for women than it does for men; that is, language skills are a tool for accruing social status. Women often have a wider range of styles than men (at all levels of linguistic structure: intonation, pronunciation, grammar, and discourse/pragmatic styles). But it is misleading to characterize women as a group as leaders in language change. The profiles of leaders crosscut social class and age, as well as gender. They also crosscut style or genre. Much current work on gender differences in language use explores what it is that unifies this disparate set of factors in order to better understand what specific linguistic variants mean to the users and how social and linguistic structure are dynamic and interrelated systems.

Miriam Meyerhoff

See also Conversation and Dialogue; Production of Language
GENES AND LINGUISTIC TONE

It is usually assumed that the language spoken by a human community is independent of the community’s genetic makeup, an assumption supported by an overwhelming amount of evidence. However, the possibility that language is influenced by its speakers’ genes cannot be ruled out a priori, and a recently discovered correlation between the geographic distribution of tone languages and two human genes seems to point to a genetically influenced bias affecting language. This entry describes this specific correlation and highlights its major implications.

Voice pitch has a variety of communicative functions. Some of these are probably universal, such as conveying information about the speaker’s sex, age, and emotional state. In many languages, including the European languages, voice pitch also conveys certain sentence-level meanings such as signaling that an utterance is a question or an exclamation; these uses of pitch are known as intonation. Some languages, however, known as tone languages, use voice pitch in ways that are comparable to the use of vowels and consonants: to distinguish words or grammatical structures. In Mandarin Chinese, for example, the sequence nian when pronounced with rising pitch means “year” and when pronounced with falling pitch means “read.” The number of tones distinguished in such languages, like the number of vowels and consonants in all languages, is variable but is normally between two and about seven. Probably a slight majority of the world’s languages are tone languages, though it is difficult to be precise in part because there are borderline cases that are difficult to classify as tonal or nontonal. What is certain, however, is that the geographical distribution of tone languages is patterned: They are predominantly found in sub-Saharan Africa, continental and insular Southeast Asia, and Central America and Amazonia.

Microcephaly represents a group of pathologies whereby the head (and brain) size is significantly smaller than the average. Several genes have been identified that, in deleterious mutated forms, cause specific types of microcephaly. Two such genes are Abnormal Spindle-like, Microcephaly-associated protein (ASPM) and microcephalin. Not all mutations of ASPM or microcephalin cause pathological brain reduction, and recently, two such nondeleterious variants have been identified, one for each gene (named the derived haplogroups and denoted in the following as ASPM-D and MCPH-D). These variants are of interest for the following reasons: Neither is uniformly distributed across the human species (e.g., both are rare in sub-Saharan Africa; both are of relatively recent origin (within the last 50,000 years); and both seem to be under positive natural selection, meaning that at least in some environments their effects confer some evolutionary advantage compared to other variants of the genes. However, these advantageous effects have yet to be identified; it is known that ASPM-D and MCPH-D do not correlate with head/brain size in normal individuals, with altruism, or with schizophrenia. It is possible that any advantageous effects are related to the cellular functions of the two genes and not to their involvement in brain growth and development.

The geographic distributions of non-tone languages and of the two derived haplogroups ASPM-D and MCPH-D tend to overlap, in the sense that those populations having high frequencies of both ASPM-D and MCPH-D tend to speak predominantly non-tone languages, those populations with low frequencies of both tend to speak tone languages, while other populations seem not to show any preference. This correlation is statistically highly significant, even after taking into account the two most important confounding factors: common history and contact. Common history means that populations descending from a common ancestor tend to share not only genes but also languages: If the common ancestor happened to speak a tone language and have low frequencies of ASPM-D and MCPH-D, all its descendants will tend to do the same, creating a spurious correlation between population genetics and language typology. As for contact, populations in geographical proximity tend to become more similar both genetically (through migration or intermarriage) and linguistically (through borrowing of words, sounds, or grammar), giving rise to another important source of spurious correlations.

Further Readings
While it is possible that despite careful controls the correlation between non-tone languages and ASPM-D and MCPH-D is just a coincidence, a plausible alternative explanation is that these two genetic variants somehow cause languages to be (non-)tonal. However, given that any normal child acquires the language of its community, irrespective of which variant of ASPM and microcephalin it has, any such causal relationship cannot be deterministic at the level of the individual. Instead, the causal relationship must take the form of a very small bias, whose effects are not manifest during everyday linguistic behavior but become visible only in language change, as language is transmitted across generations over time. The proposal is that in populations with enough individuals biased “against” tone but originally speaking a tone language, each new generation may develop a slightly simplified tone system leading in the end to a non-tone language.

This type of genetically based linguistic bias would allow a better understanding of the biological bases of language and its evolution from our prelinguistic ancestors, as well as deeper reconstructions of past languages. But probably the most important effect would be on the way we conceptualize the complex interactions between culture and biology, helping to move beyond the simplistic and misleading “nature versus nurture” debate toward an integrated view. Not only do genes influence culture, and not only does culture have an impact on genes, but genes need culture in order to have any effect in the first place.

D. Robert Ladd and Dan Dediu

See also Language Development; Music and the Evolution of Language; Music Perception

Further Readings


**GESTURE AND LANGUAGE PROCESSING**

When people talk, they often gesture with their hands—probably more than they realize. Speakers are generally aware of what they are saying. But when we gesture, we may not know that we’re gesturing, or we may be unaware of the information our gestures contain. Because co-speech gestures are prevalent but often unconscious, they can provide a window into how people think and communicate. This entry describes the basic types of hand gestures that typically accompany speech, provides an overview of their cognitive and communicative functions, and illustrates how gestures vary across languages and cultures.

**Types of Gestures**

Gestures vary in form, function, and in how they relate to language. *Emblems* have highly conventionalized forms and stable meanings, much like words in signed or spoken languages. Common emblems include waving the hand to say hello and giving the “thumbs-up” to show approval.

*Iconic* gestures depict some aspect of an object or action and are less conventionalized than emblems. Whereas emblems can be understood independent of language, iconic gestures are often ambiguous without the accompanying speech. A speaker tracing an arc in the air with her fingers could be depicting a rainbow, a dome, the hump of a camel, the flight of a soccer ball, or the rise and fall of a civilization. The latter would be an example of a *metaphoric gesture*, a special kind of iconic gesture that represents an abstract idea. Civilizations cannot literally rise or fall in space, but we talk about them as if they do and gesture accordingly.

*Deictic gestures* refer to objects or locations in physical or conceptual space and often complement deictic language. Telling the clerk at the donut shop that you want *that donut* may not yield the desired
result unless the spoken phrase is combined with a point to the pastry you have your eye on. Deictic gestures can convey aspects of a speaker’s meaning that are difficult to express in words.

*Beat gestures*, by contrast, may carry no meaning at all. Speakers frequently make simple motions with the hand or fingers, often repeated, and timed with prosodic peaks in speech. The cognitive and communicative functions of beats are not well understood. Some beats appear to add emphasis to ideas expressed in speech, others to serve discourse functions, and others to reveal the speaker’s emotional state: Fast, staccato beats may show agitation; precise beats can show determination or sincerity; large, forceful beats may show either frustration or enthusiasm. The significance of beats can only be interpreted in the context of the gesturer’s language, posture, or facial expressions.

Gestures rarely fit neatly into one category or another. Their functions blend and overlap. A gesturer might beat in the rhythm of their speech while pointing or vary the speed or size of an iconic gesture to endow it with metaphorical significance. As such, the gesture types described here should not be considered mutually exclusive. Multifunctional gestures that confound any simple typology are the rule, not the exception.

**Gesturing and Speaking**

Why do we gesture when we speak? One reason may be that gesturing helps speakers retrieve words more efficiently, particularly words with spatial meaning. Preventing people from gesturing makes their language production less fluent. Gestures supplement the meaning of speech in at least two ways, *matching* and *mismatching*. If a speaker cups her hand around an imaginary glass while saying *a glass of wine*, this constitutes a match (i.e., overlap) between the content of speech and gesture. If instead, the speaker holds her thumb and forefinger parallel, about an inch apart, to indicate that it was only a small glass of wine, this would constitute a speech-gesture mismatch, because size was not mentioned in the phrase *a glass of wine*. The term *mismatch*, as it is used by gesture researchers, does not necessarily suggest any incongruity between speech and gesture; rather, a mismatching gesture provides information that is not available in the co-occurring speech. Whether matching or mismatching, speech and gesture are never fully redundant. Speech provides a *selective description* and gesture a *selective depiction* of an idea, each highlighting certain aspects. Together, speech and gesture form a composite communicative signal.

**Gesturing and Thinking**

But is gesture only for communicating? If so, why do people still gesture when they’re on the telephone? In principle, gesturing when nobody can see us could be a vestige of gesturing during face-to-face communication. Yet communicative habits cannot explain why congenitally blind children gesture similarly to sighted children in some contexts, even though they have never seen gestures and have no experience with their communicative function.

Gestures serve cognitive functions for the speaker, independent of their impact on the listener. In classroom settings, gestures can aid learning. More generally, gestures help with tasks that require maintaining or transforming spatial and motoric information in memory.

**Cross-Linguistic Variation in Gesture**

The way information is packaged in a language’s grammar affects how its speakers gesture. For example, in languages such as English, clauses that describe motion events typically encode information about both the *manner of motion* (e.g., swinging, rolling) and the *trajectory* (e.g., down, across). In other languages, such as Turkish and Japanese, manner and trajectory are packaged into separate clauses. Gestures by speakers of these languages differ accordingly: Speakers of languages that separate manner and trajectory syntactically are more likely to express these aspects of motion events in separate gestures.

Although some emblematic gestures are recognizable across language communities, others are language specific. French and Italian speakers use the “my eye” gesture, pulling down the lower eyelid with the index finger to indicate skepticism about what someone is saying, but this action has no conventional meaning for English speakers. Other gestures have strikingly different meanings across communities. “The horns,” made by extending the pinkie and index finger while making a fist, is used to ward off the evil eye in traditional Mediterranean cultures. Variants of this gesture were used in Elizabethan England to accuse a man of having an
unfaithful wife, in modern England and the United States to express a passion for heavy metal music, and in the southern United States to show allegiance to the University of Texas Longhorns sports teams.

Are some gestures universal? NASA carried this assumption to new heights when they affixed a picture of a man showing his open palm to the Pioneer 10 spacecraft, in the hope that this gesture of friendship would be interpretable by any extraterrestrials who should find it.

Daniel Casasanto

See also Conversation and Dialogue; Multimodal Conversational Systems

Further Readings

**GROUP DECISION MAKING**

Decisions are often made in groups: sometimes to maximize the quality of the decision, sometimes because the decision makers are all stakeholders, and sometimes to diffuse, or obscure, responsibility. Group decisions range from formal committees to casual conversations, and research has examined the impact on quality, creativity, dedication, and risk tolerance. There are ways in which it is advantageous to have groups involved in decision making. However, in groups, other processes and motivations will often come into play that impair the quality of the decision making. In general, convening a group to make a decision is rarely beneficial, especially when compared to aggregating the same number of independent opinions. The benefits of interaction are not likely to outweigh the costs associated with irrelevant or even counterproductive goals that are added by group settings. There are, however, processes that can minimize such detrimental effects.

**Benefits**

Having a group make a decision has various potential benefits. Some of these are quite separate from the quality of the decision itself and include having people feel enfranchised, adding to the credibility of the outcome, and distributing responsibility for the result. However, there is also some notion that the decision itself can be better when it is made by a group. A group is able to include a greater variety of viewpoints and thus summon a wider wealth of knowledge. People’s idiosyncratic biases, prejudices, and ignorance can be canceled in the aggregate. Another potential benefit is that people will, in a group, be inspired by those around them, increasing their motivation to do well and seeding their own creativity with the ideas of others. Brainstorming procedures are designed to take advantage of the social interactions, with each person’s ideas inspired by and building on the suggestions of others, and the outcome cumulated over the wisdom of each participant. There is evidence that aggregated opinions can be remarkably good in what is sometimes called the *wisdom-of-crowds-effect*, but the way the aggregation occurs is critical in order to avoid the potentially powerful negative effects of group processes.

**Social Comparison Effects**

One effect of making a decision as a group is that the individuals do not operate independently of each other, but are motivated to compare their positions to others’ and to modify their behavior to manage the impression they make on others. These social comparison processes can have various effects on the decision. One factor that it can alter is the level of risk that is tolerated. This finding, generally called the *risky shift* or, more accurately, the *group polarization phenomenon*, suggests that group discussion will shift the level of risk generally in the direction that is admired. Thus, if the decision concerns sports, for example, where people are inclined to take risks and admire those who do, the group decision will support greater risk than did the average individual. On the other hand, if it concerns the well-being of children, where caution is preferred, the group will adopt a safer strategy than would individuals. At least part of the explanation for this phenomenon and for it working in both directions is that people wish to be slightly better than the average person,
although also not very different from the rest of the group. In the sports scenario, then, discovering during the group discussion that one is actually at about the group average would cause one to slide toward being a bit braver, and so the whole group would move in that direction.

Social comparison processes can impact group decisions in ways other than risk tolerance, and many major decisions that were not only demonstrably wrong but also should have been identifiable as wrong at the time they were made have been ascribed to the operation of group processes. There are several ways that group processes contribute to such faulty decision making. One is that people's desire to remain in the group can prevent them from voicing negative impressions of possible decisions. This self-censoring of concerns will be especially strong when membership in the group is very precious, as is often the case with plum political appointments. Another process is that people will judge the appropriateness of voicing concerns by comparison with the other members of the group. Thus, there can arise an impression of unanimous support, even though each member secretly harbors grave doubts.

**Motivational Effects**

In addition to allowing people to respond to each other's behavior and modify their own position as a result, a group can also change the motivation and efficiency of its members. The start of this work goes back over a century to investigations of the effects on performance of the presence of others. In some cases, performance is enhanced, but in others, it is impaired, and the dividing factors appear to be the complexity of the task and whether individual or collective group output is identifiable. For many sorts of group decisions, both of these factors would work to prevent optimal outcomes, as the decisions are often hard and the outcome collective. For many brainstorming situations, for example, which also tend to have difficult tasks and do not identify individual contributions, the evidence suggests that the net result is to reduce the number of ideas and their quality. Part of this is due to coordination losses—forgetting one's idea while others are talking. Part is due to the fear of embarrassment. But a good part is due to some diffusion of responsibility; with others available to take up the slack, and one's own potential efforts divided between too many, the effort does not seem warranted. For these reasons, many group settings lead not to greater efficiency but to social loafing.

**Overcoming the Limitations**

There are ways to reap the benefits of group decision making while reducing the costs of social comparison biases and motivational deficits. It is often possible to aggregate independent positions rather than have group discussions and thus keep the benefits of independence or uncorrelated error. Sharing people's opinions or input and then having people reconsider, again as individuals, can allow for cross-pollination and inspiration while still limiting negative social comparison effects. Some biases can also be overcome by the appointment of a devil's advocate, and it has been shown that the breaking of the unanimous façade by a member is effective even when that member's position is not one with which others agree. It is only necessary that the person demonstrate that not all agree and thus that it is possible to express that disagreement. Biases can also be reduced by reducing the insularity in the group and by allowing in outside voices for whom the motivation to remain in the group is not present and the norm of implicit unanimity is not established. The motivational effects can be limited by making individuals responsible for particular outcomes rather than allowing responsibility to be diffused across the group.

_Nicholas J. S. Christenfeld_

See also Dissent, Effects on Group Decisions; Social Loafing; Wisdom of Crowds Effect

**Further Readings**


GUILT

In contrast to the socio-legal meaning of guilt, which refers to the attribution of culpability regarding the commission of a blameworthy act, most emotion scholars view guilt as an adaptive psychological process that promotes social cohesion. This entry begins with a discussion of how the term guilt is defined by scholars who study guilt as a psychological process. The empirical research literature on this concept is then briefly reviewed.

Guilt as a Self-Conscious, Social-Moral Emotion

For social scientists who study the mind, guilt is defined as a self-conscious, social-moral emotion consisting of an unpleasant affective state, often accompanied by thoughts about the self having engaged in a blameworthy violation of a social norm (e.g., John felt guilty about not returning Mary’s phone call). Guilt also refers to instances in which an individual experiences unpleasant feeling states while merely contemplating the future violation of a social convention or rule, in which case we refer to this as “anticipated” guilt (e.g., Mary felt guilty when she thought about not leaving a tip for the waitress).

Guilt is a social-moral emotion in that it is intimately related to social welfare, in the sense that guilt is triggered by real or perceived violations of culturally valued conventions and rules. As such, guilt is part of a family of emotions known as “moral sentiments,” which serve as a sort of “social glue,” allowing social relations to function efficiently. According to 18th-century economist Adam Smith, moral sentiments such as guilt serve this social function by virtue of their capacity to compel individuals to do one of the following: (a) abide by social conventions (i.e., an ostensible function of guilt) or (b) to generate applause toward others who succeed (i.e., an ostensible function of the moral sentiment admiration) or approbation toward those who fail (i.e., an ostensible function of the moral sentiment contempt or reproach) to uphold such standards.

Because the experience of guilt entails a focus on one’s own adherence to social norms and conventions, guilt is seen as a self-conscious emotion rather than an other-focused emotion. Although other social-moral emotions such as contempt or reproach entail unpleasant feelings and thoughts generated in response to the blameworthy actions of others (e.g., John felt angry because Mary did not return his phone call), the emotion guilt always entails a focus on the actions (or inactions) of the self. Even in the case of “collective guilt,” whereby one incurs an unpleasant feeling while considering the blameworthy actions of a group with which one strongly identifies, the focus is ultimately on the self (e.g., John felt guilty when he learned that his firm was responsible for polluting the river).

Distinguishing Guilt and Shame

In everyday language, guilt is often confused with another self-conscious, social-moral emotion, shame. Guilt can be distinguished from shame in terms of the focus on evaluating one’s own actions or omissions (i.e., guilt) versus evaluating one’s personhood or self (i.e., shame). In this light, the term guilt more properly refers to a negative evaluation of one’s own behavior (e.g., feeling guilty because you accidentally bumped into another car in the parking lot), whereas shame refers to a negative evaluation of one’s personhood (e.g., feeling ashamed because you believe that you are a bad person). Another important distinction between guilt and shame centers on the contrast between the sometimes adaptive consequences associated with experiencing guilt and the often maladaptive correlates of shame proneness. Research by June Tangney and others has shown that guilt is an adaptive emotion that facilitates cooperation, whereas shame is often seen as a less adaptive emotion, promoting withdrawal or externalizing in the form of aggression. The next section briefly reviews the research on the social-behavioral consequences of guilt.

Behavioral Effects of Guilt

Because guilt can be experienced in situations where an individual is merely contemplating the future violation of a social norm, some economists have argued that the unpleasant affective state that accompanies guilt can serve as a commitment device. By referring to guilt as a commitment device, economists such as Jack Hirshleifer have argued that this psychological state (e.g., feeling guilty) serves as a powerful incentive to remove one’s unpleasant feeling state
by acting to repair the damage to social relations caused by one’s previous (or potential) actions or inactions. For example, when guilt occurs after the performance of an action that is perceived to have damaged some aspect of one’s social relations, this psychological state can serve as a powerful incentive to remove one’s unpleasant feeling state by performing some act of compensation directed toward the offended party. Several studies have found that inducing an irrelevant state of “guilty feelings” can compel individuals to avail themselves of the next available opportunity to cooperate with another. One early study showed that inducing individuals to accidentally break the experimenter’s camera caused participants to be subsequently more inclined to help an unrelated third party. More recent research suggests that the tendency for guilty feelings to motivate helping behavior also occurs under more natural circumstances in which feelings of guilt are endogenous to the situation at hand. These findings suggest that guilty feelings may have their strongest impact on individuals with preexisting prosocial motives, the very same individuals who are least likely to engage in noncooperative behavior in the first place. Another focus of empirical research on guilt has been on “guilt proneness,” or the tendency to experience this emotion across a range of situations as a stable feature of one’s disposition. In this regard, guilt-prone individuals have been found to be less likely to respond aggressively when angered, compared to less guilt-prone individuals. Consistent with the ostensive prosocial functions of guilt, these studies reveal that guilt-prone individuals tend to experience more empathy toward others, which appears to serve as an important mediator to the link between guilt proneness and inhibition of aggression. In sum, guilt proneness, unlike shame proneness, appears to be associated with positive impacts on social functioning. Although there has been a trend toward an increasing focus on more pernicious moral sentiments such as anger and disgust, research on guilt continues to represent about 15% of all research on moral emotions.

Timothy Ketelaar

See also Emotion and Moral Judgment; Rationality of Emotion; Self-Consciousness

Further Readings


The study of happiness has long been a playground for philosophical speculation. Because of a lack of empirical measures of happiness, it was not possible in the past to check propositions about the matter. In the late 20th century, survey research methods introduced by the social sciences brought a breakthrough. Dependable measures of happiness were developed, by means of which a significant body of knowledge has evolved. This entry presents an overview of that knowledge and discusses (a) the concept of happiness, (b) the measurement of happiness, (c) the prevalence of happiness, (d) the conditions for happiness, and (e) the promotion of happiness.

What Is Happiness?

The word *happiness* is used in many meanings. The different meanings are presented in the schemes below.

**Four Qualities of Life**

Quality-of-life concepts can be sorted using two distinctions, which together provide a fourfold matrix. The first distinction is between chances and outcomes, that is, the difference between opportunities for a good life and the good life itself. A second difference is between outer and inner qualities of life, in other words, between external and internal features. In the first case the quality is in the environment; in the latter, it is in the individual.

<table>
<thead>
<tr>
<th>Outer Qualities</th>
<th>Inner Qualities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Life chances</td>
<td>Livability of</td>
</tr>
<tr>
<td></td>
<td>environment</td>
</tr>
<tr>
<td>Life results</td>
<td>Utility of life</td>
</tr>
<tr>
<td></td>
<td>Satisfaction</td>
</tr>
</tbody>
</table>

**Livability of the Environment**

This phrase denotes the quality of living conditions. Different disciplines emphasize different aspects of the environment. Natural conditions are emphasized by ecologists who describe livability in terms of pollution, global warming, and the degradation of nature. The built environment is emphasized by city planners who associate livability with such things as sewer systems, traffic jams, and ghetto formation. Society is central in the sociological view on livability, both the quality of society as a whole and the relative position one has in society. Livability is not the same as what is called happiness in this entry. It is a generic term for environmental preconditions for happiness, and not all these are equally conducive to happiness.

**Life Ability of the Person**

The next phrase denotes how well we are equipped to cope with the problems of life. Individual capability involves in the first place absence of functional defects—that is, *health* in the limited sense, sometimes referred to as *negative health*. A second aspect is optimal functioning, commonly referred to as *positive health* and associated with energy and resilience. In that line, capability is also placed in a
developmental perspective and seen to include acquisition of new skills for living. The term self-actualization is often used in that context. Like livability, life ability is not the same as what is called happiness in this entry. If one is competent in living, one has a good chance at happiness, but this endowment does not guarantee an enjoyable outcome, and not all capabilities are equally conducive to happiness.

Utility of Life

The bottom left phrase represents the notion that the quality of a life is not only in that life itself but also in the contribution it makes to other things. This is sometimes referred to as the “meaning” of life. Different external effects are considered in the context of utility: the impact a life has on other people, its contribution to human civilization, or even its long-term effects on the biosphere. In moral philosophy, life is judged on its moral or esthetic value, and exemplary lives are seen as better than standard lives. Such “virtuous living” is often presented as the essence of “true” happiness but is not the same as happiness as defined in this entry.

Core Meaning: Subjective Enjoyment of Life

Finally, the bottom right term represents the inner outcomes of life; that is, the quality in the eye of the beholder. As we deal with conscious humans, this quality boils down to subjective enjoyment of life. This is commonly referred to by terms such as subjective well-being, life satisfaction, and happiness, in a limited sense of the word. This is the kind of happiness addressed in this entry.

Four Kinds of Satisfaction

Even when we focus on subjective satisfaction with life, there are still different meanings associated with the word happiness. These meanings can also be charted in a fourfold matrix. In this case, that classification is based on the following dichotomies: life aspects versus life as a whole and passing delight versus enduring satisfaction.

<table>
<thead>
<tr>
<th></th>
<th>Passing</th>
<th>Enduring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Part of life</td>
<td>Pleasure</td>
<td>Domain satisfaction</td>
</tr>
<tr>
<td>Life as a whole</td>
<td>Top experience</td>
<td>Life satisfaction (happiness)</td>
</tr>
</tbody>
</table>

Pleasure

The top left table cell represents passing enjoyments of life aspects. Examples would be the joy of a sunset, the delight of a good meal, or the appreciation of a concert. This kind of satisfaction concerns hedonic experiences, both sensory pleasure and mental enjoyments. This is not quite the same as what is called happiness in this entry, which denotes a broader meaning and concerns both overall satisfaction and life as a whole. Pleasure contributes to a positive appreciation of life, but it is not the whole of it.

Domain Satisfaction

The top right phrase denotes more lasting appreciation of life aspects, such as satisfaction with one's family life and satisfaction with work. Though satisfaction with such domains of life depends typically on a flow of pleasures, the appraisals have some continuity of their own. For instance, one can remain satisfied with one's children even if one has not enjoyed their company for some time. Sometimes the term happy is used in the context of domain satisfactions, such as when we speak about a happy marriage or being happy with a job. In this entry the term happiness is used in the broader sense of satisfaction with life as a whole. We would not call a person happy who is satisfied with work and family but still dissatisfied on the whole because of bad health.

Top Experience

The bottom left denotes ecstatic moments that involve the perception of wholeness. In these experiences intense pleasure goes together with a positive view on life as a whole. This is the kind of happiness poets write about. Again this is not the kind of happiness addressed in this entry. A moment of bliss is not lasting enjoyment of life. Actually such momentary peaks tend to harm long-term satisfaction, because among other things, they distort the view of reality.

Lasting Satisfaction With One's Life as a Whole

The bottom right represents the combination of enduring satisfaction and satisfaction with life as a whole. This is what is meant by the word happiness in this entry. A synonym is life satisfaction. Happiness in this sense is defined as the “degree to
which an individual judges the overall quality of his or her own life as a whole favorably”—in other words, how much one likes the life one leads. Ed Diener defines subjective well-being (SWB) in much the same way.

**How Do We Assess How Happy We Are?**

In evaluating their life, people draw on two sources of information: (a) how well they feel most of the time and (b) to what extent they perceive that their life meets current standards of the good life. These subappraisals are referred to as components of happiness—respectively, the affective and cognitive component. The affective component is called hedonic level of affect. It is commonly assumed that affective experience draws on the gratification of needs, which are part of human nature. The cognitive component is called contentment and is seen to depend on culturally variable wants. The available data suggest that the affective component dominates in the overall evaluation of life, which implies that happiness depends more on needs than on wants.

**Measurement of Happiness**

Since happiness is something we have in mind, it can be measured using single direct questions. An example of a survey question on overall happiness is the following:

Taking all together, how satisfied or dissatisfied are you currently with your life as a whole?

0 1 2 3 4 5 6 7 8 9 10
Dissatisfied Satisfied

Components of happiness can be measured in this way. Hedonic level of affect can further be measured using multimoment assessment—for instance, the experience sampling method (ESM), in which people who have been given beepers record how happy they feel at the moment they are beeped. All measures of happiness are listed in the collection *Measures of Happiness* in the World Database of Happiness and linked to research findings obtained with these.

**How Happy Are We?**

In 2008, the average response to the above question was 7.0 in the United States. The highest score was observed in Denmark (8.4) and the lowest in Zimbabwe (3.3). The world average is about 6.0, so most people are happy. Still not everybody is equally happy. For example, 13% of Americans rated 5.0 or lower, while 16% ticked 10. All research findings of this kind are in the collection *Happiness in Nations* in the World Database of Happiness.

**What Determines Happiness?**

Most of these differences in average happiness across nations are due to the quality of society. Not surprisingly, people live happier in nations that provide a good material standard of living, safety, freedom, and justice. What may be a surprise is that people are happier in modern, individualistic societies rather than in traditional, collectivistic societies and that average happiness is not lower in nations where income disparities are great. Together, these societal characteristics explain about 75% of the observed differences. Social conditions for human happiness are fairly universal.

Social factors explain less of the differences in happiness within modern western societies. Only some 10% can be attributed to income, education, and social rank. Some 15% seems to be due to strokes of good or bad luck, while about 30% is due to genetic makeup. A big deal of the difference seems to be in learned art-of-living skills, such as social intelligence. The new “positive psychology” aims at identifying these aptitudes and finding ways to enhance them. Research results are summarized in the collection *Correlational Findings*, part of the World Database of Happiness.

**Can Happiness Be Fostered?**

Some believe that happiness is relative and that chasing after it will get you as far as a mouse in a treadmill. Others say that happiness is a fixed trait and as such is practically unchangeable. Research shows, however, that happiness can indeed be raised lastingly. Average happiness has gone up in most of the contemporary nations over the last 40 years, though long-term follow-up studies have shown that we do not adapt to everything—for example, the loss of a child.

**Should Happiness Be Fostered?**

For some, happiness is the greatest good and one should aim at greater happiness for the greater
number of people. Many religions see this differently and place more value on human suffering. Research into facts cannot determine whether enjoying life is morally better than suffering from it. Research statistics do offer some insight into the consequences of viewpoints and show to what extent seeking happiness meshes with other values. In this connection, research was carried out into the extent to which happiness brings out the good or the bad in people. It appears that happiness does not breed contented cows but rather activates people. Happiness broadens our scope and helps build up resources. Another striking result is that happiness is good for your health and that happy people live longer. Happy people are better citizens; they need fewer scapegoats, give more of themselves for social organizations, and are, perhaps, more sensible voters. In short, fostering happiness gives more than just a more pleasant life. In a number of ways, subjective happiness can make life objectively better as well.

How Can Happiness Be Fostered?

Happiness can be fostered at three levels: (1) At the macro level, happiness depends heavily on the quality of society. There are wide differences in average happiness across nations, and these differences are clearly linked to societal characteristics, such as economic prosperity and political freedom. Policymakers can do a lot about that. Still, not everything that policymakers are concerned about does add to happiness; for instance, income inequality in nations does not influence happiness. (2) At the meso level, happiness depends on the institutional settings in which we spend most of our time, such as in work and school. Systematic improvements in those realms will probably add to the happiness of a great number of people. (3) At the micro level of individuals, happiness can be fostered in several ways. One way is to strengthen abilities to cope with the problems of life. Investments in education and therapy are required for that purpose. Another way is informing people about the typical consequences of major life choices for happiness, such as early retirement. This requires investment in long-term follow-up studies, such as the Happiness Monitor Project.

Ruut Veenhoven

Further Readings


HEARING

Hearing probably initially evolved to alert organisms to significant events in the environment. Unlike visual stimuli, sounds can be detected whatever their direction of incidence. Hearing can indicate the presence of a predator or prey, and it can indicate the appropriate location to direct visual attention. Hearing can also convey information about the properties of sound-emitting objects, such as their size and whether they are approaching or receding. In complex acoustic environments, when multiple sound sources are active, the auditory system is usually able to analyze the complex mixture so as to derive a percept of each of the individual sound sources. In humans, hearing has evolved further to allow speech communication. The ability to convey information using sounds as symbols for objects and actions is one of the characteristics that distinguishes humans from other animals. This entry covers the following topics:

See also Affective Forecasting; Emotion, Cultural Perspectives; Emotion, Structural Approaches; Self-Knowledge
• The analysis and representation of sounds in the auditory system
• The relative roles of analysis at peripheral and central levels of the auditory system
• Processes involved in the masking of one sound by other sounds
• Perceptual interpretations and auditory illusions
• How the auditory system analyzes mixtures of sounds to derive percepts corresponding to individual sound sources
• How the perception of auditory objects can remain stable despite changes in the conditions of listening

How Information About Sound Is Carried in the Auditory System

The basic structure of the early stages of the auditory system is illustrated in Figure 1. Sounds are transmitted through the outer ear (the pinna and ear canal, or meatus) and middle ear (which includes three very small bones, called the malleus, incus, and stapes, collectively known as the ossicles) into the inner ear, which includes the cochlea. Within the spiral-shaped cochlea, there is a kind of ribbon, called the basilar membrane, which runs from the tip of the spiral (the apex) to the outer end of the spiral (the base). The ribbon is surrounded by fluids. When a sound enters the ear, the basilar membrane moves up and down. Each place on the basilar membrane is tuned to respond best to a limited range of frequencies. Low frequencies produce their biggest response toward the apex of the cochlea, and high frequencies produce their biggest response toward the base.

Lying on top of the basilar membrane are specialized sensory cells called hair cells. These run in rows along the length of the basilar membrane. One type of hair cell, called the inner hair cell, responds to the movement of the basilar membrane by generating an electrical signal that in turn leads to the release of a neurotransmitter that triggers activity in the neurons of the auditory nerve. Each neuron derives its response from the vibration at a specific place on the basilar membrane.

Figure 1  Schematic illustration of the structure of the peripheral auditory system, showing the outer, middle, and inner ear

Information about the characteristics of sounds is carried in the auditory nerve in three basic ways:

1. By the rate of firing of individual neurons, which will be referred to as the “amount” of neural activity. The more vibration there is at a given place, the greater is the amount of activity in neurons connected to that place. It is commonly believed that the subjective loudness of a sound is related to the amount of neural activity evoked by that sound, although this idea has been disputed.

2. By the distribution of activity across neurons. Each neuron is tuned so that it responds most strongly to a specific frequency, called the characteristic frequency (CF); the tuning reflects the tuning of the place on the basilar membrane that drives the neuron. The distribution of the amount of neural activity as a function of CF is called the excitation pattern. The excitation pattern conveys “place” information since the CF at the peak of the excitation pattern is related to the place on the basilar membrane that is excited most.

3. By the detailed time pattern of the neural impulses and especially the time intervals between successive nerve impulses. This form of information is known as “temporal” information. Neural impulses tend to be evoked at times corresponding to a specific phase of the waveform on the basilar membrane (for example, at the peaks of the waveform), an effect called phase locking. As a result, for a periodic sound, the time intervals between successive nerve impulses are approximately integer multiples of the period of the sound. For example, if the sound has a frequency of 500 hertz (Hz), the period is 2 milliseconds (ms), and the time intervals between successive nerve impulses would be close to 2, 4, 6, 8, 10, . . . ms. Phase locking breaks down at high frequencies (above about 36 kHz [kilohertz] in most mammals), but the upper limit in humans is not definitely known. Studies of pitch perception suggest that phase locking is very weak for frequencies above about 5 kHz.

In addition, information about sounds is conveyed by the differences between the two ears in all the above. In particular, differences in intensity at the two ears (primarily conveyed by differences in neural firing rate) and differences in the time of arrival of sounds at the two ears (conveyed mainly by subtle differences in the exact timing of nerve spikes) play a strong role in determining the perceived location of sounds.

**Place Versus Temporal Information**

A classic debate in hearing theory is concerned with the relative importance of place and temporal information. Originally, this debate revolved mainly around the relative role of place and temporal information in the perception of the pitch of pure and complex tones. The debate about pitch continues to this day, but most researchers now believe that pitch perception involves the use of both place and temporal information; indeed, it may be the case that there has to be a correspondence between the two types of information for a clear pitch to be perceived. More recently, researchers have proposed that both place and temporal information play a role in several other aspects of auditory perception, including masking (the process whereby one sound is rendered inaudible by the presence of another sound) and speech perception.

**Peripheral Versus Central Processes**

Peripheral processes in hearing are those occurring in the outer and middle ear, in the cochlea, and in the auditory nerve. Central processes are those occurring at stages of the auditory system following the auditory nerve. Many theories and models of auditory perception are based on relatively peripheral processes. For example, models of loudness perception take into account the effect of the outer and middle ear on the sound transmitted to the cochlea as well as the processing of sound within the cochlea. However, the processing that occurs in the central auditory system has been largely ignored. Early theories of pitch perception were also largely based on the information that was present in the auditory nerve. Only in the case of sound localization, which requires comparison of the neural signals from the two ears, were central processes considered. The trend over the last 20 years has been to consider the role of central processes in much more detail, both at the physiological and the psychological level.
Energetic and Informational Masking

An example of the increasing consideration of central processes comes from studies of auditory masking, which is the process by which one sound (the masker) makes it difficult or impossible to hear another sound (the signal or target). Many of the cases of masking studied in the laboratory, for example, the detection of sinusoidal tones in white noise, can be explained largely in terms of processes occurring in the cochlea and auditory nerve. There are models of masking that can predict the detection thresholds of typical human listeners for such cases with high accuracy. Masking of this type is sometimes called “energetic masking,” and it occurs when the response of the auditory nerve to the masker-plus-signal is very similar to the response to the masker alone. In other words, the signal is masked because the information conveyed by the auditory nerve is inadequate for its detection.

When a masking sound is highly similar in some way to the signal and/or when the properties of the masker vary in an unpredictable way from one stimulus to the next, there may be much more masking than would be expected from energetic masking alone. This “extra” masking is called informational masking. It is assumed that informational masking occurs because the signal is confused with the masker, or because attention is directed to an inappropriate aspect of the sound.

When the task of a listener is to identify the speech of one talker (the target) in the presence of another talker (the background) with similar characteristics, informational masking may play a strong role. Under these conditions, the amount of informational masking is greatly influenced by whether or not the target and background are perceived to come from the same location in space. Informational masking is reduced by a difference in perceived location of the target and background.

Informational masking, and the conditions under which it occurs, has been an area of increasing research activity over the last 20 years. One finding of this research is that there are large individual differences; some people appear to be much more susceptible to informational masking than others. There can also be considerable learning effects, with informational masking decreasing over many weeks of daily training. The existence of such large learning effects confirms that the masking is not determined by peripheral processes but depends on limitations in central mechanisms, which can partly be overcome by training.

Veridical Perception and Auditory Illusions

Many researchers in the field of perception take the view that our sensory systems are designed to help us determine the true nature of the outside world. When such veridical perception fails, the percept is often described as an illusion. For vision, it is usually relatively obvious what the “correct” perception ought to be; for example, a straight line ought to be perceived as straight. For hearing, it is often much less obvious how a sound ought to be perceived, and it is often not clear whether an auditory percept should be described as an illusion.

Consider as an example the perception of the pitch of a complex tone. Tones produced by musical instruments are usually composed of a series of sinusoidal components, called harmonics. For example, for a specific note the frequencies of the harmonics might be 200, 400, 600, 800, 1,000, . . . Hz. The lowest harmonic (200 Hz) is referred to as the fundamental component. The complex tone is heard as having the same pitch as a single sine wave (also called a pure tone) with a frequency of 200 Hz. However, if the fundamental component is removed from the complex tone, the pitch is not altered. In other words, the pitch corresponds to the fundamental frequency, even when the sound contains no component at the fundamental frequency. This is called the phenomenon of the missing fundamental. In the early days of research on pitch, this effect was sometimes described as an illusion. Nowadays, the phenomenon is simply assumed to reflect the way that pitch perception normally works. The pitch of a complex tone is usually determined by components (higher harmonics) other than the fundamental frequency. It became inappropriate to refer to the phenomenon as an illusion once better theories of pitch had been developed.

As another example, consider the so-called continuity illusion. When a Sound A is alternated with a Sound B, and B is more intense than A, then A may be heard as continuous, even though it is interrupted. For perceived continuity to occur, the gaps in A must be filled by the Sound B, and B must be a potential
masker of A (if they were presented simultaneously). Although the continuity effect is often described as an illusion, the auditory system’s interpretation of A as being continuous is perfectly reasonable. If A were continuous, the pattern of neural activity evoked by A plus B would not be detectably different from the pattern evoked by the alternation of A and B. Generally, the auditory system seems to create a percept that corresponds to the most likely or most plausible interpretation of the sensory information. It is thus not clear whether it is appropriate to describe the continuity effect as an illusion.

Auditory Scene Analysis and the Concept of Auditory Events and Objects

It is hardly ever the case that the sound reaching our ears comes from a single source. Rather, there are often several sources active simultaneously. The auditory system is faced with the task of analyzing the complex mixture and deriving a perceptual representation of each individual sound source. The task is immensely difficult and complex, and most computational methods that have been developed to perform the task—often described using the term computational auditory scene analysis—perform much more poorly than human listeners.

Albert Bregman has emphasized the distinction between two concepts: source and stream. A sound source is some physical entity that gives rise to acoustic waves—for example, a person talking. A stream is the percept of a group of successive and/or simultaneous sound elements as a coherent whole, appearing to come from a single source. For example, it is the percept of hearing a person talking. Some researchers have argued that the auditory system has evolved primarily to allow it to accomplish the task of deriving streams when multiple sources are active. Millions of years of evolution have led to the superiority of human listeners over machines in performing this task.

Perceptual Constancy

Perceptual constancy refers to the fact that the perceived properties of objects remain relatively constant despite changes in the conditions of viewing or listening. For example, the shape of a coin does not appear to change when it is viewed from an oblique angle, and the perceived color of a person’s face does not change markedly when viewed in different conditions of lighting, despite the fact that the wavelength of the light being reflected from the face may be dramatically different. Perceptual constancy has been studied intensively for the visual system but has received relatively little attention for the auditory system. Consider the case of loudness perception. The physical intensity of the sound reaching a listener’s ears depends partly on the sound intensity emitted by the source and partly on the distance of the listener from the source. It remains somewhat unclear whether loudness is related more to the intensity at the ears or to the intensity of the sound emitted by the source. However, at least when the sound source is visible, a form of constancy sometimes holds; the intensity at the source plays a stronger role than the intensity at the ears. In other words, the auditory system appears to base loudness mainly on the properties of the source, the distal stimuli, and not the properties of the sound reaching the ears, the proximal stimulus. For example, when the sound intensity at the ears is held constant, judgments of loudness are influenced by the perceived distance of the source, as determined visually.

Perceptual constancy may also play a role in the perception of the timbre of sounds, which refers to their characteristic quality. If two tones have the same loudness and pitch but are perceived as different, then the difference is described as a difference in timbre. Timbre depends partly on the shape of the spectrum of sounds. In many everyday situations, the sound reaching the ears is composed partly of sound coming directly from the source and partly of sound that has been reflected from nearby surfaces and reaches the ears after a delay. The interference of the direct and reflected sound changes the spectrum of the sound at the ears relative to the sound emitted by the source. The influence of the reflected sound has been studied extensively in terms of its effects on sound localization, which are surprisingly small. However, the effect on timbre perception has received little attention. One might expect the reflected sound to have a strong influence on perceived timbre, but everyday experience suggests that this is not the case. Rather, perceptual constancy seems to operate. For example, the sound quality of a familiar person’s voice does not seem to differ markedly when listening to them outdoors and in a reverberant room. The studies of Anthony Watkins and colleagues suggest that listeners are
able to compensate for the effects of room acoustics (reflections from the walls, floor, and ceiling of a room), at least for the purpose of identifying speech sounds. However, further work in this area is clearly needed.

Brian C. J. Moore

See also Auditory Masking; Audition, Neural Basis; Music Perception; Perceptual Constancy; Speech Perception

Further Readings


**HEARING, PHILOSOPHICAL PERSPECTIVES**

Hearing and auditory perception are rapidly developing topics in the philosophy of perception. Recent work has focused on characterizing what we hear and on similarities and differences between audition and other modalities. Future work should address how theorizing about audition impacts theorizing about perception more generally.

This entry concerns questions about the objects and contents of hearing. It includes discussion of the spatial content of audition, of the role of time and pitch in the individuation of auditory objects, and of audition’s role in the perception of speech.

**Objects of Hearing**

One sort of question about what we hear concerns the objects of auditory perception. What kinds of things do we hear? We hear sounds. But what kinds of things are sounds? Recent answers include the following: sensible properties, events, or something that depends on pressure waves. We can also ask what other sorts of things we hear. For instance, we might auditorily perceive things and happenings that make sounds, such as clarinets, crashes, or conversations. Perhaps we even hear silence.

**Contents of Hearing**

A related question about what we hear concerns the contents of auditory experience. In hearing, how do we experience the world to be? One common way to pose this question appeals to the correctness or veridity conditions for auditory perceptual experiences. Framed in this way, audition’s content is prima facie relevant to theorizing about audition’s objects.

**Hearing and Space**

One central philosophical question about auditory content concerns space. Hearing furnishes information about space: Thanks to hearing, you can learn that the fridge is to the left or that the dump truck is far away. Does audition represent space or have spatial content? An extreme view is that it does not. Some claim that although we “work out” spatial facts on the basis of audition, hearing itself is aspatial; it reveals no spatial features.
However, empirical research on spatial hearing suggests that audition does have spatial content. Subjects auditorily discern direction and distance in a way that suggests perception rather than “working it out.”

Nevertheless, vision and audition differ with respect to space. Following P. F. Strawson’s example, some philosophers hold that audition, unlike vision, is not intrinsically or inherently spatial. On such an account, audition may have some spatial content, but it inherits that content from another spatial modality, such as vision or tactile-kinesthetic perception. So a purely or exclusively auditory experience would not be spatial. Perhaps this is because audition lacks vision’s inherent spatial structure.

An alternative account rejects that audition is inherently aspatial. The difference between vision and audition, instead, is said to be one of degree. Vision’s spatial content is more fine-grained and accurate than audition’s.

Furthermore, the spatial characteristics of the objects of vision and audition differ. Vision’s objects include cohesive three-dimensional items akin to material objects. Not only do they seem located in space relative to each other, but they also have a rich internal spatial structure. While audition’s objects include individual items located in three-dimensional space, they do not seem to have the rich internal spatial structure of vision’s objects. Vision’s objects are perceptually individuated and identified in terms of their spatial features, while space is less important to individuating and identifying audition’s objects.

**Hearing, Time, and Pitch**

How are audition’s objects individuated and identified? Two kinds of features are most important. First, time plays a role in audition analogous to space in vision. Sounds are temporally extended, and they are individuated and identified in terms of their temporal structure. The sound of a spoken word, a melody, or a police siren is individuated in part by its patterns of change in time.

Second, pitch also plays a role in individuating audition’s objects. Distinct pitches from a single source (such as a loudspeaker) frequently are heard as two distinct audible individuals. Tones of the same pitch from different sources frequently are heard as one audible individual, as when two loudspeakers play the same pitch. In audition, unlike vision, spatial discontinuity may be neither necessary nor sufficient for distinct audible individuals.

Some philosophers have argued on these grounds that auditory objects are complex individuals with temporal rather than spatial structure. Others, in part on the basis of pitch, have challenged whether the role of time in audition is analogous to that of space in vision.

**Speech and Hearing**

No discussion of hearing is complete without considering speech. Three questions are of philosophical interest.

First, what accounts for the phenomenological difference between listening to speech in a language you know and in one you do not know? The difference might be just that in the former case you grasp or understanding the meanings associated with the sounds you hear. The phenomenological difference thus is entirely cognitive—it stems from the phenomenology of understanding. Or the difference might be at least partly perceptual. One might, for instance, hear meanings or auditorily represent semantic properties. Or one might hear language-specific attributes such as phonological features. On an austere perceptual account, just the audible acoustic features of the sounds of speech might differ once you know a language.

Second, what are speech perception’s objects? The traditional account is that hearing speech involves hearing sounds, then grasping their associated conventional meanings. Speech perception and ordinary nonlinguistic audition thus share objects. But empirical evidence indicates that speech sounds, such as phonemes, do not map in a consistent way onto features of the acoustic signal. Some authors therefore reject the view that speech perception and ordinary audition share objects. Antirealists argue that apparent speech sounds are mere mental constructs. Others suggest that, unlike ordinary audition, speech perception’s objects include articulatory gestures of the mouth and vocal tract. This realist account is bolstered by the fact that visual cues about mouth gestures impact the perception of speech, as shown by the famous McGurk effect, in which a visual presentation of a moving mouth alters how a voice sounds.

Third, to what extent is the capacity to perceive speech distinct from ordinary audition? Some hold that speech perception involves a dedicated
perceptual modality. Others hold that speech perception is a variety of ordinary audition, whose objects and perceptual resources are continuous with ordinary hearing. This requires an account of the special significance of speech sounds for humans—even neonates prefer speech to nonspeech sounds.

Casey O’Callaghan

See also Hearing; Naïve Realism; Smell, Philosophical Perspectives; Speech Perception

Further Readings


Heritability

The concept of heritability comes to us from population genetics. The idea was developed but not named by the geneticists Sewall Wright and Ronald Fisher. Both were theorists and applied their ideas to both agricultural and human psychological traits (i.e., intelligence). Wright developed the method of path analysis, which is now widely used in behavior genetics to compute heritability coefficients. These methods are very flexible and have been modified to incorporate modern molecular genetic findings. This entry outlines the purpose of the heritability formula, describes its various components, provides some examples of how the heritability of a trait is computed, and discusses some of the limitations of the concept (what it does not tell us).

The heritability coefficient indexes genetic influence on a quantitative trait (i.e., height, weight, intelligence) for a specific population. It tells us how much of the variance (a measure of the extent of differences between people) in a population is due to genetic factors. It is a population statistic (describes a population, just as the mean of a trait is descriptive of a population) and does not apply directly to individuals. It is useful to distinguish between “broad heritability” and “narrow heritability.”

Broad heritability tells us the proportion of total phenotypic (measurable) variance ($\sigma^2_p$) in a trait that is due to all genetic influences ($\sigma^2_g$). It can vary from sample to sample, environment to environment, and age to age. Thus the formula:

$$h^2_{\text{broad sense}} = \frac{\sigma^2_g}{\sigma^2_p}.$$ 

This formula should be read as broad heritability = all genetic variance/phenotypic variance.

The term $h^2$ is the symbol for heritability and is not to be interpreted to mean the value is to be squared.

We know from Mendelian theory that genes act in a number of ways. Many act additively: The more genes positive for the trait the more the trait expresses itself (height is a good example). The idea is that these additive genes, each of small effect, sum up to influence a trait. In addition, some genes dominate other genes (Mendelian dominance). This is a form of interaction (nonadditivity) between two genes at the same place (locus) on the paired human chromosomes. Another form of nonadditivity is epistasis, the interaction between genes at different loci on the chromosomes. Additive genes create similarity between relatives in a straightforward manner. Nonadditive effects work in a much more complex way and their effects are difficult to estimate with precision in human studies.

Narrow heritability includes only the proportion of variance because of genes that act in an additive manner. This is called the additive genetic variance ($\sigma^2_a$). Thus the formula:

$$h^2_{\text{narrow sense}} = \frac{\sigma^2_a}{\sigma^2_p}.$$ 

As in the previous example, this formula should be read as narrow heritability = additive genetic variance/phenotypic variance. Again the term $h^2$ is the symbol for heritability and is not to be interpreted to mean the value is to be squared.

Environmental sources of variance are, of course, conceptually important. One such source is called shared environmental variance or common
environmental variance ($\sigma^2_c$). This environmental influence is what has been thought to make twins and siblings similar (similar child rearing, same home, parents with the same educational background, etc.). For a long time psychologists thought that common environment was an extremely important source of twin and sibling similarity, but behavior genetic studies have shown that it is of minor importance for many traits measured in adulthood. For a number of traits (i.e., intelligence and social attitudes), shared environment is very important in the early years but much less so in adulthood. We can estimate the importance of shared environment in a number of ways, but two are particularly simple. Both use correlation coefficients. The correlation coefficient used in twin and adoption studies can run from zero to one and can, in the example below, be interpreted as a percentage of variance. In most other contexts, a correlation coefficient must be squared to interpret the results as a percentage. This point is widely misunderstood. Consider monozygotic twins reared together (MZT), often incorrectly called identical twins. They have all their genes in common. Consequently, the correlation between them on a trait such as intelligence reflects all the effects of genes (additive and nonadditive) plus the similarity because of being reared in the same family (shared environment). Monozygotic twins reared apart (MZA), if they were placed in adoptive homes randomly, would be similar only because they had all their genes in common (additive and nonadditive effects). Thus the MZA correlation would directly estimate the broad heritability. The difference between the two correlations (MZT – MZA) would estimate shared environmental influence. The IQ correlation between MZT twins, in adulthood, is about .86. The IQ correlation for MZA twins (almost all adults) is .74 (weighted mean of five studies). The difference suggests a shared environmental influence of .12. The second method of estimating shared environmental influence is to examine the similarity of unrelated individuals reared in the same home (URT). For childhood data the URT correlation is .26 (weighted mean of 14 studies). In adulthood the URT correlation is .04 (weighted mean of five studies). This method suggests a shared environmental influence of .04. Other methods yield estimates between these two figures and also show a strong age effect, with genetic influence increasing with age.

If the heritability of adult IQ is .74 (there is disagreement about this value and figures range from .50–.85) depending on varying assumptions, the common environment is .12, and the reliability of IQ tests is about .90, then the remaining (nonshared) environmental influence is about .04.

For some traits, there are sources of variance that depend on both genetic and environmental factors taken together. These include genotype × environment interactions ($\sigma^2_{ge}$) and genotype by environment correlations ($\sigma^2_{g*e}$). These sources of variance are difficult to study, but with new genetic tools becoming available almost daily, it is likely they will prove to be an important addition to our understanding of the various sources of variance, and they may well change current views of how genes and environment transact and interact.

It is important to realize that the methods discussed above give us a picture of the effects of genes and environments in terms of variance in the traits studied. This is a common method in psychology where analysis of variance is used extensively. It is, however, a particular level of analysis; there are other levels of analysis. Heritability does not tell us, for example, about how genes or the environment actually work. It is also important to realize that the expression of a trait can increase in a population even though the heritability of the trait is quite high. Most of the variation in stature is genetic, but there has been a continuing increase in stature in industrialized populations for many years. There has been a similar increase in the raw scores on IQ tests, widely known as the Flynn Effect.

Thomas J. Bouchard Jr.

See also Flynn Effect: Rising Intelligence Scores; Intelligence, Neural Basis

Further Readings
Johnson, W., Turkheimer, E., Gottesman, I. I., & Bouchard, T. J., Jr. (2009). Beyond heritability: Twin studies in

**Heritage Language and Second Language Learning**

In a bilingual context, an individual is exposed to and learns two or more languages. Different types of bilingualism are defined by the sociopolitical status of the languages (majority vs. minority language), the order of acquisition of the languages (first vs. second), and the degree of exposure to and use of each language in different contexts throughout the life span. Second and heritage language learning are just two examples.

**Second Language Learning**

Second language learning typically refers to the acquisition of a second language after the structural foundations of the first or native language are in place. Second language learning can occur in childhood between the ages of 4 and 12 or in adulthood (around and after puberty). Immigrant children who arrive in a host country are considered child second language learners. For example, immigrant children in the United States are second language learners, whereas English-speaking children receiving second language instruction a few hours a week at school are foreign language learners. The amount and quality of input, the reasons for learning the language, as well as the opportunity to use the language on a daily basis differ in a foreign language versus a second language context, when the second language can be the majority language spoken in the wider speech community. Immigrant children must learn the second language to thrive socially and academically in the new society. By contrast, for majority language-speaking children, learning a foreign language is a choice. Adult second language acquisition can also occur in a second language environment, again in the case of immigrants, or in a foreign language context through schooling.

The field of second language acquisition, which has evolved considerably since the 1960s, is generally concerned with understanding the linguistic development of the second language from initial state to ultimate attainment in child and adult learners, the internal and external factors that play a role in the process, and the ways input can be manipulated through instruction. Much research in this field has been guided by drawing comparisons with the field of first language acquisition. It is typically assumed that monolingual children eventually became native speakers of the language and achieve full linguistic competence in both the structural (phonology, morphology, syntax, semantics) and the communicative functions of their language (contextually appropriate). However, the degree of success in the second language is typically variable: Not every second language learner achieves native-like knowledge and fluency. Age of acquisition (before and after puberty), cognitive development, type/amount of input, and affective factors, among others, affect the eventual outcome of the learning process. While children learning their first language in a monolingual context overcome typical grammatical errors, many second language learners do not, displaying what Michael Long has termed *fossilization* or *stabilization*. Some persistent errors in second language acquisition are caused by transfer or interference from the first language at the cognitive level.

**Heritage Language Learning**

If the language of the host country is the second language in children of immigrant families, their family language is the heritage language. Heritage language learners are bilingual children born in the host country or children of immigrants who speak a minority language at home and learn a second language in the wider community and at school. The emerging field of heritage language acquisition is concerned with language maintenance and loss at the individual level and with how to teach speakers who possess different degrees of proficiency in their heritage language. Like many adult second language learners, adult heritage language speakers do not reach native-like attainment in their heritage language. In fact, the range of variability spans from mere receptive knowledge to fully fluent speakers with advanced literacy skills and all the shades in between. Among
the possible reasons for why heritage speakers fail to develop their minority language, Silvina Montrul and Maria Polinsky have proposed that reduced exposure to and use of the language during childhood and beyond leads to incomplete acquisition and to attrition of their native language. Incomplete acquisition refers to the fact that a given feature of the language, which is present in the linguistic repertoire of native speakers of that particular linguistic variety, is not mastered at age-appropriate levels. For example, many adult Spanish heritage speakers make gender and verbal agreement errors typical at age 2 or 3 in Spanish-speaking children but are not common in adult speakers. It is possible that agreement was not fully mastered by age 4, when most children cease to make these errors in Spanish. Attrition, on the other hand, implies that a given structure reached a certain degree of mastery but was subsequently lost. If a child had learned the future tense at age 5, at the same level as monolingual children of the same age, but showed very high error rates 2 years later, then one can establish through a longitudinal research design that attrition took place.

Thus, as immigrant children go through the process of assimilation to mainstream society and learn their second language, their heritage language learning is often severely compromised. Many heritage language speakers develop full linguistic competence in their second language (especially if they were born or arrived in the host country in early childhood), but linguistic development in their heritage language lags significantly behind. In fact, although their heritage language learning started in childhood—a key factor in successful language learning—their linguistic development, the types of grammatical errors they make, and the general outcome of the linguistic process have much more in common with second language acquisition than with first language acquisition. At the same time, by virtue of language learning in a naturalistic setting, many heritage language learners lack the highly developed metalinguistic and literacy skills of instructed second language learners.

_Silvina Montrul_

**Further Readings**


**Human Classification Learning**

Human classification learning is the process of learning to distinguish between different types or groups of things in the world. Humans are often presented with various situations in which people (friend or foe), objects (edible or toxic), and places (safe or dangerous) must be correctly classified to survive. Classification allows us to respond to these situations appropriately and further allows these situations to be identified and incorporated into groups, categories, or concepts based on similar properties. This entry will further define human classification learning as well as highlight different ways in which classification can be studied. Finally, the entry will briefly introduce the neural architecture underlying human classification learning.

To group or categorize various types of objects and situations, learners can either learn the group membership of each object or situation individually or can generalize by finding common attributes within a group of objects, with the aim of being able to apply the same action to these distinct, yet similar, objects or situations in future encounters. Classification learning can have several uses. Classification allows us to bypass the process of full identification of a stimulus, to quickly generalize across objects or situations in our environment on the basis of simple dimensions, and to respond appropriately to the environment. Classification allows us to quickly and efficiently distinguish among a wide range of objects even without the benefit of prior experience. In addition, classification enables us to
easily group different objects so that more complex grouping structures within specific categories can be abstracted. In other words, classification is the initial process by which objects are parsed into separate groups before category membership can be determined, a necessary step one must go through to access information acquired through category learning.

Though similar, human classification learning should not be confused with category learning or even concept learning. Overall, classification emphasizes the processes involved in assigning stimuli to groups, whereas concept learning emphasizes the mental representations of our semantic knowledge of the world. Categorization often combines the process of classification with the mental representations of concept learning. However, at times these terms have been used interchangeably, so researchers and students in this area should take care when reading the scientific literature.

Experimental Approaches

Because of its wide scope, human classification learning can include any task that requires the learner to make distinctions between the group membership of two or more stimuli. Despite this, classification is typically studied experimentally using three approaches. One approach is known as trial and error learning, or feedback-based learning. In a typical study, the goal of a learner is to produce an appropriate response for any stimulus without the benefit of previous training examples. To achieve this, the learner must view the stimulus, then make an initial response, and finally receive feedback. Based on the resulting feedback, the learner must abstract sorting criteria and generalize these rules to new stimuli in some “reasonable” way. Eventually, this trial and error method will enable the learner to produce the correct response more and more frequently. A second approach is instructed learning, or paired-associate learning. Learners are initially told how each stimulus should be classified, with the goal of eventually being able to classify stimuli without being given the answer. A third approach is known as free sorting. This approach requires learners to partition stimuli along one or more dimensions without information about whether their response was correct or incorrect. Learning has to rely on the distinguishing attributes of the stimuli for the stimuli to be classified. Formally, trial and error learning and instructed learning are both types of supervised learning, whereas free sorting is a form of unsupervised learning; the theoretical differences between these two approaches are beyond the scope of this entry.

Neural Substrates of Classification

During the past decade our knowledge regarding the neural substrates of human classification has grown by leaps and bounds. Recent research suggests that classification learning relies heavily on the basal ganglia, particularly the striatum. Because of its location, the striatum receives and integrates input from the majority of the cortex. In addition, the striatum sends information back to the cortex, making the striatum essential for many different cognitive processes, including classification. However, the striatum is not a homogeneous structure that plays a single role in classification. Rather, the striatum is comprised of subsections that each plays a distinct role in human classification.

Learning to classify, especially by feedback, can be broken down into several different steps thought to depend on different functional regions of the striatum. Studies indicate that the nucleus accumbens, or ventral striatum, is essential for the detection of feedback as part of the reward system. Similarly, the head of the caudate nucleus in conjunction with prefrontal cortex is involved in integrating the executive functions involved in classification with feedback and reward-related information and possible motor responses. Furthermore, research shows that the body and tail of the caudate and the putamen are implicated in learning to correctly classify stimuli: the body and tail of the caudate in conjunction with temporal lobe cortex may subserve perceptual aspects of classification, whereas the putamen in conjunction with motor cortex may be involved in selecting and executing motor behaviors appropriate for the stimulus's class. The midbrain dopamine system interacts with the striatum via ascending and descending loops that connect striatal regions in a hierarchical fashion and support coordination of these separate functions that subserve classification learning.

Besides the striatum, various neural areas make significant contributions to classification depending
Hypochondria

This entry provides a brief overview of the concept of hypochondria and its historical background; this disorder plays an important role in psychological and somatic health care systems. The entry also discusses recent theoretical models, which have led to an empirically supported series of psychological treatments, including those using a cognitive-behavioral frame of reference.

Nature and Origin of Hypochondriasis

Hypochondria (also written as hypochondriasis) is characterized by unjustified fears or convictions that one has a serious and often fatal disease, such as heart disease, cancer, or AIDS. The main component is health anxiety, and in contemporary writing this more neutral term is often preferred over hypochondria(sis). Patients frequently seek reassurance in medical consultation, check their bodies, and avoid illness-related triggers (such as bodily sensations and medical information). Merely informing the patient of the absence of a disease process or explaining the benign nature of the symptoms results in temporary reassurance followed by renewed worry over symptoms and continuing overuse of medical services. Hypochondria is thought to be quite common, and a number of studies estimate its prevalence in the general population as between 4% to 7%, with equal numbers for men and women. The course is often chronic, but remission (also known as transient hypochondriasis) also occurs, and it has been found that about one in three patients no longer meet criteria for hypochondriasis after a 5-year follow-up period.

Traditionally, hypochondriasis (an ancient Greek word meaning below the cartilage) was considered to be a special form of melancholia (i.e., depressed mood) resulting from an excess of black bile. In the 17th century, scientists argued that hypochondriasis occurred only in men and was equivalent to hysteria occurring in females. Around the turn of the 20th century, the Austrian psychoanalyst Sigmund Freud suggested that hypochondria was the consequence of an imbalance in sexual energy that would build up and result in physical symptoms. Other early psychological theories on hypochondriasis suggested that physical symptoms developed in individuals defending against low self-esteem because a sick body is attached to fewer stigmas than a sick mind. These early theories were overly speculative, lacking empirical support from research.

Theoretical Perspectives

At the end of the 20th century, more specific cognitive-behavioral models were formulated in which great emphasis was placed on the patient’s misinterpretation of innocuous bodily sensations. A number
of studies on the information processing features of hypochondria supported these models. It was shown that selective attention to illness information, risk perception, misinterpretation of benign symptoms, and cognitive responses to medical reassurance are maintaining factors. It has also been found that hypochondriacal individuals hold distinct (catastrophic) assumptions about health and illness. However, it has not been convincingly demonstrated that triggering these assumptions (e.g., by exposure to illness related stimuli) leads to increased hypochondriacal concerns. Several studies found a substantial positive correlation between hypochondria and the tendency to experience a broad range of bodily sensations as noxious, intense, and disturbing. Far less attention has been paid to specific conditioning models regarding the onset and maintenance of hypochondriasis, although theories of excessive health anxiety stipulate that internal cues (e.g., pain sensations or stomachache) perceived as predictors of threat and bodily harm (unconditioned stimuli—i.e., naturally threatening circumstances) can serve as conditioned stimuli (i.e., originally neutral stimuli that acquire the meaning of the naturally threatening stimuli). In the presence of these stimuli, individuals will exhibit so-called conditioned responses—that is, learned reactions. In hypochondriasis these responses are generally related to anxiety and its concomitant physiological sensations (e.g., muscle tension, palpitations).

**Onset and Maintenance**

To date, a comprehensive and empirically tested model of the onset and course of hypochondriasis is lacking, although the cognitive behavioral hypothesis is generally adopted as an important heuristic. In this conceptualization of hypochondriasis, the disorder is assumed to originate from threatening situations (e.g., witnessing the death of a loved one) that form maladaptive health-related core assumptions (i.e., “My body is vulnerable”). These assumptions are triggered under circumstances in which ambiguous bodily sensations (e.g., palpitations) are experienced or in the event of critical incidents (such as being confronted with somebody else’s symptoms or even death). Next, the assumptions may be consolidated by further health-related experiences, by selective attention to illness related information, and by confirmatory reasoning bias (i.e., only attending to support of the hypochondriacal assumptions). It is generally believed that the age of onset for hypochondriasis lies in early adulthood, although, in individual patients, any age may be the start of the disorder. A few studies have concentrated on a genetic and familial component in hypochondriasis, suggesting that there is, at best, only a moderate genetic contribution. Several researchers have suggested that exposure to disease in the family—in combination with parental attitudes toward illness—influences the development of hypochondriacal concerns in children. It has also been found that an insecure (fearful) attachment style is positively related to hypochondriasis. Finally, sexual traumatization also has been linked to the development of excessive health anxiety. In a sample of hospital outpatients, hypochondriacal adults recalled more childhood trauma (parental upheaval, sexual trauma, and victimization by violence) before the age of 17 than a control group of patients.

**Treatments**

Until the late 1980s, psychological and psychiatric treatments for hypochondria were generally considered hardly feasible. Some authors went as far as stating that there was no specific treatment and that patients would not profit from any treatment whatsoever. In the 1990s, research into the nature and treatment of hypochondriasis increased. In particular, the cognitive behavioral conceptualization of the disorder has stimulated the development of new treatments. In particular, cognitions and behaviors are therapeutic targets, and even attentional processes seem promising candidates in this respect. This has resulted in an increasing number of treatment studies in which the clinical effectiveness of cognitive behavior therapy has been established. Psychopharmacological treatment has been little studied, although some authors note positive results.

_Theo K. Bouman_

*See also* Behavioral Therapy; Reinforcement Learning, Psychological Perspectives
Further Readings


Idealism

Idealists believe that reality is fundamentally mental in some sense, or that the physical world is somehow dependent on, or constituted by, the operation of minds. Idealism is the polar opposite of materialism, which can be understood to be the view that ultimate reality is wholly material and that mentality is dependent on material processes.

This entry discusses three forms of idealism: the idealism of the 18th-century philosopher George Berkeley, the phenomenalism of the 20th century, and panpsychism.

Berkeley

The 18th-century philosopher George Berkeley is probably the most famous idealist in the history of Western philosophy. Berkeley believed in the existence of only two kinds of thing: minds and ideas. Ideas are dependent on minds for their existence: For an idea to exist is for it to be perceived by some mind. Although Berkeley denied the existence of material objects, understood as mind-independent entities, he claimed to accept the commonsense view that there are ordinary objects, such as tables, chairs, rocks, and planets. However, for Berkeley, such objects, which he was happy to call “physical” objects, are entirely constituted of ideas. A red, juicy apple is composed of the red idea we perceive when we look at it, the idea of solidity we perceive when we touch it, and the idea of juiciness we perceive when we bite into it (note that Berkeley, in keeping with the empiricist tradition of his time, uses the word idea to mean something like “a mental thing which is the immediate object of sensory perception”).

Part of our commonsense understanding of physical objects is that they remain in existence even when no one is around to look at them. Berkeley had two strategies for accommodating this ordinary belief. The first strategy is that the apple exists by virtue of the fact that if someone were to look in its direction, she or he would perceive the visual ideas that make it up (and if she were to touch it, she would perceive the idea of solidity that makes it up, and if he were to bite into it, he would perceive the idea of juiciness which makes it up, etc.). The second strategy is that the apple exists by virtue of the fact that, even in the absence of human perceivers, God is permanently perceiving the apple: seeing its redness, feeling its hardness, tasting its juiciness.

It is crucial in understanding Berkeley’s view to see him in relation to John Locke, the British empiricist who came before Berkeley and whose life overlapped with Berkeley’s. Locke claimed that the immediate objects of perception are sensory ideas (call this claim A), although he also believed that we indirectly perceive mind-independent material objects. Locke argued that some qualities of objects, such as color and taste, are dependent for their existence on the mind, on the grounds that our ideas of these qualities vary according to the perspective of the perceiver (call this argument B). However, he also claimed that other properties, such as solidity and extension, are mind independent.
Berkeley took claim A and argument B to what many believe to be their logical conclusion. If the only things we immediately perceive are ideas, it is difficult to see how we could ever have evidence for the existence of mind-independent objects. If the fact that ideas of color and taste vary depending on the perspective of the perceiver gives us grounds to think that color and taste are mind dependent, then we also have grounds for thinking that solidity and extension are mind dependent, as the ideas of these qualities also vary depending on the perspective of the perceiver (the idea of a object’s shape varies depending on the angle from which we look at it).

Of course, one way around this motivation for idealism is simply to deny the assumptions that Berkeley shared with Locke (indeed, many philosophers reinterpret Berkeley’s arguments as *reductio ad absurdum* style arguments against these Lockean assumptions). However, Berkeley also has arguments for idealism that are independent of those assumptions. Berkeley argues that a mind-independent world is inconceivable and therefore not possible (in the same way that square circles are inconceivable and therefore not possible). Try to imagine something existing without the mind, a tree that no one is looking at, for example. But surely you yourself are perceiving this tree in your imagination, and hence the tree is not in fact an object that is not perceived. Try as you might, an object that is not being perceived cannot be conceived of.

Few philosophers accept this argument, but it is difficult to see where it goes wrong. One suggestion is that Berkeley is confusing the distinction between properties the imagined tree has or lacks, and properties I am imagining the tree to have or lack (which may be contrary to the properties the imagined tree in fact has or lacks). It may be the case that the tree has the property of being perceived (due to the fact that I am imagining it), but it does not follow that I cannot imagine the tree lacking that property. Barack Obama has the property of being the president of the United States, but his having of that property is consistent with my imagining him lacking that property—for example, if I imagine that he is a schoolteacher. In the same way, just because any object I imagine is thereby being perceived (by me in my imagination), it does not follow that I cannot imagine it not being perceived.

A common error that undergraduate students make when studying Berkeley is to think that Berkeley’s arguments for idealism rely on the existence of God. This is to get things precisely the wrong way round. Berkeley argues for idealism independently of his belief in God and then uses idealism as a premise of an interesting argument for the existence of God. Whenever you open your eyes, you find pressed upon you ideas that you yourself do not cause—that is, the ideas of physical objects in the world around you. But what causes these ideas? Given idealism, it cannot be that they are caused by material objects, for there are no such things. Berkeley believes in only two kinds of things: ideas and minds. Ideas, for Berkeley, are essentially passive and can do nothing but be perceived by minds. It follows that the thing that causes the ideas in our minds of physical objects must be itself a mind. Given the complexity and sophistication of the physical world, this mind must be a very powerful, intelligent mind. Given the beauty of nature, and its helpfulness to humanity, it is natural to suppose that this mind is beneficent. We quickly find the cause of our ideas having many of the traditional qualities of the personal God of the Judeo-Christian-Muslim tradition.

**Phenomenalism**

The 20th-century position, *phenomenalism*, a view advocated by A. J. Ayer, among others, is very similar to Berkeley’s idealism. The phenomenalists believed that material objects are constituted by sense data, and by *sense data* they meant something very similar to what Berkeley meant by *ideas*. They also held, and again the similarity with Berkeley’s view is clear, that the red, juicy apple exists in the absence of a human mind on account of the fact that, if someone looked in its direction, he or she would have a red sense datum, and if someone bit into it, he or she would perceive a juicy sense datum.

However, whereas Berkeley grounded these conditional truths in the will of God, the phenomenalists took them to be brute facts about reality, not capable of being explained in more fundamental terms. This aspect of phenomenalism was later strongly criticized by C. B. Martin, on the grounds that it entails that there are truths without “truth-makers”—that is, truths about the world that aren’t made true by anything in the world. For Berkeley, the true sentence, “If I bite into the apple, it will taste juicy” is made true by the will of God, but for
the phenomenalist this sentence is just plain true, even though there is nothing in the world that makes it true. (This strategy of criticizing various metaphysical positions for failing to provide truthmakers has subsequently become very popular in analytic philosophy. For example, some people attack presentists, i.e., philosophers who don’t believe in the reality of the past or future, on the grounds that they are unable to provide truthmakers for [what most presentists would agree are] true sentences about the past, e.g., “Dinosaurs used to roam the Earth.”)

Panpsychism

Berkeley and the phenomenalists take the physical world to be fundamentally mental in the sense that for a physical object like a table, a planet, or an electron to exist is for it to be perceived, or for it to be perceivable. The philosophical position known as panpsychism offers a different way of taking the world to be fundamentally mental in character. The panpsychist agrees with the materialist that physical objects have an existence entirely independent of their being perceived. In what sense then, is physical reality fundamentally mental for the panpsychist? The panpsychist believes that the intrinsic nature of fundamental physical objects is, at least in part, mental. Whereas a materialist (at least of a standard kind—some panpsychists take their view to be itself a form of materialism) holds that the intrinsic nature of an electron is constituted entirely by nonmental properties (e.g., spin, mass, and charge), the panpsychist holds that the intrinsic nature of the electron is partly constituted by mental properties. In this sense the fundamental nature of reality is, at least in part, mental.

It is easy to caricature the panpsychist position, for example, by seeing it as holding that tables feel jealous or that beer cans suffer from existential angst. However, it is important to bear in mind two important qualifications the panpsychist is likely to emphasize. First, the panpsychist need not hold that macroscopic inanimate objects such as tables and chairs are bearers of mental properties. They are more likely to suppose that only fundamental physical particles and certain metaphysically significant composite objects, such as organisms or their brains, directly instantiate mental properties. Tables and beer cans are intrinsically mental only in the sense that the fundamental particles that compose them have mental properties. Second, the panpsychist need not hold that the mental properties of fundamental physical particles are intentional properties; she need not hold that a fundamental particle has thoughts. She is more likely to claim that an electron has some very simple kind of raw experience, a kind of experience massively simpler than the complex and sophisticated experience that organisms enjoy.

Despite these qualifications, panpsychism is still seen by many as a somewhat eccentric position. However, it has enjoyed a renaissance of late, perhaps most influentially in the work of Galen Strawson, attracting followers with its promise of offering a way of giving a theoretically satisfying explanation of the existence of the consciousness of organisms. On a conventional materialist view of things, fundamental particles are in no sense conscious, and yet, in certain very specific arrangements, they compose composite objects—organisms or the brains of organisms—which do have consciousness. Many philosophers worry that this alleged emergence of experience from the wholly non-experiential, seems like nothing short of a miracle and gives us a rather disunified and theoretically dissatisfying picture of nature. Panpsychism offers us a unified view of the natural world that avoids this whiff of the miraculous. For the panpsychist, just as the tiny mass of fundamental particles in the organism constitute the greater mass of the organism, so the very crude forms of consciousness instantiated by these same particles constitute the consciousness of that organism (or of its brain).

Idealism continues to be a minority view. There seems to be a basic repugnance in our intellectual culture to taking mentality to be among the fundamental furniture of the universe. This repugnance is perhaps most vividly expressed by Samuel Johnson in his symbolic rejection of Berkeley: “I refute in thus!” he said, kicking a stone (it is not clear how this constituted a rejection of Berkeley, given that Berkeley believed in stones). But there are also strong arguments for idealism and theoretical advantages to adopting the view. It is for these reasons that idealism has always remained a serious, if minority, metaphysical picture of the world.

Philip Goff

See also Consciousness and Embodiment; Eliminative Materialism; Physicalism; Reductive Physicalism
Implicit Memory

The term memory is typically used to refer to conscious recollection of past events, a state in which current mental contents are recognized as a product of prior experience. However, researchers have long supposed that memory for the past can influence present behavior without conscious recollection. The Swiss neurologist Édouard Claparède provided a famous example in his study of a patient with Korsakoff’s syndrome, a disease resulting in amnesia. This patient had great difficulty remembering new events that occurred after the onset of the disease. Claparède suspected, however, that the patient had residual memory abilities that did not produce conscious recollection. To test this possibility, Claparède hid a pin in his hand and gave the patient a painful pinprick on shaking hands. When Claparède next met the patient, she refused to take his offered hand. When asked for an explanation, the patient could not provide a reason for her refusal to shake hands. When pressed, she said that people sometimes hide pins in their hands as a type of practical joke. Clearly the patient formed some type of memory of the original event, but this memory was not recollected as a personal experience. Hermann Ebbinghaus, in his pioneering analysis of memory, similarly argued that memories for past events could influence behavior without conscious recollection, a view that prompted the development of the savings measure of memory.

Modern research in psychology, neuropsychology, and neuroscience supports a distinction between conscious and unconscious influences of memory, embodied in the distinction between explicit and implicit memory. Explicit memory refers to intentional or conscious recollection of prior experiences. Implicit memory, in contrast, refers to influences of the past that are not accompanied by intentional or conscious recollection. This entry describes how these two forms of memory differ from one another. It also describes the major theories accounting for these differences, and reviews recent neuroimaging studies that delineate the brain regions underlying explicit and implicit memory.

Research on Implicit Memory in the Modern Era

Research on anterograde amnesia provided the major impetus for modern research on this topic. This is the form of amnesia exhibited by Claparède’s patient. It has traditionally been defined as a decreased ability to learn and retain information about new experiences coupled with otherwise normal perceptual and intellectual abilities. Research in the 1980s and 1990s complicated the traditional definition of amnesia and demonstrated surprising, preserved memory ability, reminiscent of Claparède’s demonstration. A classic study of this type was reported by Elizabeth Warrington and Lawrence Weiskrantz. In this study, amnesics and normal control subjects were presented with a series of words. Later, their memory for the words was tested either explicitly, by asking them to recall the words, or implicitly. In the latter case, a fragmented or partial word (e.g., MET___) was presented for completion (e.g., METAL). For this test, participants are not required to recall or consciously remember any of the studied information. However, memory for the prior words can be expressed in an increased tendency to complete the fragments with previously studied words. On the explicit recall test, the amnesics exhibited the defining symptom of their disorder: They recalled many fewer words than did the control subjects. On the word completion test, both groups were more likely to complete fragments with previously studied words, and this improvement occurred to the same degree. Thus, the amnesic subjects showed the same
level of retention as the normal control subjects did on the implicit test, despite dramatically deficient conscious recollection.

**Dissociations Between Implicit and Explicit Memory**

The foregoing experiment presents the standard approach to studying implicit memory. The typical experiment consists of a study phase followed at some delay by a memory test. In the study phase, subjects are presented with a carefully controlled experience, such as a list of words or pictures presented on a computer screen. The memory test assesses how much of the studied material is retained. In standard explicit tests, subjects are directed to think back about the studied material. Examples include free recall, in which subjects report all that they can remember about the studied material in the absence of any overt cues, and recognition, in which subjects try to differentiate studied stimuli from new material. Implicit tests of memory, in contrast, do not direct subjects to think back. Rather, these tests are presented as separate exercises, unrelated to previous components of the experiment. Implicit tests are often presented as perceptual tasks, general knowledge tests, or problem-solving tasks. However, the tasks are carefully designed to exhibit influences of the recently experienced study materials. For example, the word-stem and word-fragment completion tests are typically presented as language fluency tests, in which parts of words are presented and subjects complete them with the first word that comes to mind (e.g., ban____ or b _ n _ n _ for banana). Subjects are unaware that some of the stems and fragments correspond to studied words and others correspond to new words that were not presented during the study phase. The proportion of completions is typically greater for the studied words than for the new words. This difference can occur even in the absence of intention to retrieve the studied materials and in the absence of awareness that the implicit task relates to the earlier portion of the experiment. Other implicit tests use conceptual as opposed to fragmentary perceptual cues. For example, on the category-exemplar generation task, participants are presented with the names of taxonomic categories (e.g., FRUIT) and are asked to generate examples from each. This task is presented as a general knowledge or semantic fluency test, but some of the categories correspond to words on the study list and others do not. Subjects are more likely to produce studied examples than counterbalanced, non-studied examples. In general, the increased accuracy, ease, or speed of processing for studied versus non-studied materials is referred to as priming, and it serves as the measure of implicit memory. Although priming experiments are the primary measure of implicit memory, this topic has also been studied with measures of sequence learning and classical conditioning.

The principles that govern implicit and explicit memory differ in several ways, as demonstrated by dissociations between implicit and explicit tests. One of the most striking dissociations is exemplified by the studies showing that anterograde amnesics perform poorly on explicit memory tests but normally on implicit tests. This indicates that organic amnesia impairs conscious recollection of the past but not unintentional, unconscious manifestations of memory. Similar dissociations exist in which populations impaired in explicit memory exhibit normal levels of implicit memory. For example, patients with schizophrenia or depression often demonstrate decreased explicit memory coupled with normal performance on implicit measures of memory. Likewise, healthy older adults usually have reduced explicit memory ability compared to healthy young adults, but the two age groups often perform equivalently on implicit tests of memory.

Experimental variables also produce dissociations between implicit and explicit memory, providing additional evidence for the separability of the two forms of memory and insight into their functional differences. For example, manipulations of attention during encoding can produce different effects on implicit and explicit memory. Experiments examining attention and memory typically present study materials to some subjects under full attention conditions, while other subjects encode study materials while simultaneously carrying out a secondary task, such as monitoring a sequence of tones and categorizing each as high or low. The secondary task is designed to distract subjects, reducing their attention to the study stimuli. And, of course, it works: Dividing attention during encoding reduces performance on recall and recognition tests compared to the full attention condition. However, if memory is tested with implicit tests, such as word-fragment completion or lexical decision, dividing attention
Implicit Memory

produces little effect; the two groups produce equivalent priming. Thus, dividing attention at encoding can dissociate performance on explicit and implicit tests. A similar dissociation is produced by varying the amount of semantic processing during the study episode, which has a marked effect on explicit memory tests (the levels-of-processing effect) but typically has little effect on such implicit memory tests as word-stem completion and perceptual identification.

Manipulations of perceptual attributes of stimuli often affect priming but not explicit memory. For example, word-fragment completion and word-stem completion produce more priming for study words that had been presented visually as opposed to aurally. In contrast, study modality typically produces little effect on explicit tests like free recall and recognition. It should be noted that the critical issue is the match in modality between study and test. Although most implicit tests are presented visually, priming tasks can be implemented in the auditory modality as well. For example, auditory word stems or fragments can be presented in the auditory version of the word-stem and word-fragment tasks. These auditory tasks show a similar sensitivity to study modality, although in this case, the aurally presented study words exhibit more priming than visually presented words.

Some experimental variables primarily influence explicit but not implicit memory (e.g., divided attention, semantic elaboration), whereas others primarily influence implicit but not explicit (e.g., study modality). Still other variables produce opposite results on implicit and explicit tests. For example, study items presented as pictures produce better recall and recognition (i.e., better explicit memory) than study items presented as words; this is known as the picture superiority effect. On implicit tests such as word-stem and word-fragment completion, the opposite is found: Words produce more priming than pictures. Similarly, the well-known generation effect can be reversed on implicit tests. Specifically, generating a word from a conceptual clue (e.g., hot—c___) usually produces better explicit memory than reading the same word (e.g., hot—cold). On word-stem and word-fragment completion, the opposite result obtains: Reading produces greater priming than generating. Finally, dissociations can also be found between different types of implicit tests. For example, conceptual implicit tests (like the category-exemplar production test) often produce different results than implicit tests that use perceptual cues, such as the word-fragment completion.

Multiple Memory Systems and Neuroimaging

Neuropsychological and neuroscientific data indicate that human memory is not a single ability but rather several distinct systems. One familiar distinction is that between short-term (or working) memory and long-term memory. However, research on implicit memory indicates that long-term memory is likewise not a single system. Consider the dissociations found with amnesic patients. Because anterograde amnesia typically involves damage to the hippocampus and medial temporal lobes, it appears that these parts of the brain are critical in the formation of memories that are later experienced as conscious recollection.

Early theories of multiple memory systems theories proposed dichotomies to account for explicit and implicit memory (e.g., declarative vs. procedural, episodic vs. semantic). However, it is now clear that there are several dissociable forms of implicit memory. Current theories propose four long-term memory systems: episodic memory, semantic memory, the perceptual representation system, and procedural memory. Episodic memory stores information about episodes in our personal past, enabling recollection. Semantic memory refers to our general knowledge about the world, including factual, conceptual, and linguistic information. The perceptual representation system is a perceptual memory system that processes information about the form and structure of words and objects. Finally, procedural memory represents knowledge of cognitive and motor skills. Within this framework, the episodic system is thought to give rise to explicit memory, whereas the other systems produce the unconscious, non-recollective effects of memory that constitute implicit memory.

Neuroimaging research, using positron emission tomography and functional magnetic resonance imaging, provides convergent evidence for the existence of multiple memory systems. Neuroimaging techniques allow researchers to examine the areas of the brain that are involved in implicit and explicit memory. This research has found several differences in brain activity during explicit and implicit retrieval tasks. Among the hallmarks of explicit memory retrieval is increased neural activity in prefrontal cortex (especially in the right hemisphere) and in the
medial temporal lobe and hippocampus. The activity in the frontal lobe appears to reflect a mental set in which the individual is oriented toward the past and intentionally tries to retrieve information. Activity in medial-temporal areas reflects the recollective experience itself, when the memory is successfully retrieved. Priming on implicit tests is associated with decreased activity in various brain areas. This is believed to reflect a reduction in processing demands when a stimulus is processed a second time, which in turn produces increased speed and accuracy on the priming task. The brain regions involved depend on whether the implicit test is perceptual or conceptual. For perceptual tests such as word-stem or word-fragment completion, the decreased processing is found in visual cortex (in posterior occipital lobe). In contrast, when items are re-processed on conceptual implicit tests, decreased activity is found in inferior frontal lobe and mid-temporal lobe. In general, these results imply that the same neural substrates responsible for the initial (perceptual or conceptual) processing are re-engaged at the time of test and exhibit the effects of the initial processing by their subsequent reduced activity.

Neil W. Mulligan

See also Amnesia; Consciousness and the Unconscious; Divided Attention and Memory; Memory, Neural Basis

Further Readings


Inattentional Blindness

Contrary to the popular intuition that “seeing” is simply a matter of using one’s eyes, people often fail to notice salient objects and events despite looking right at them. Inattentional blindness refers to instances when this occurs because a person’s attention was preoccupied or otherwise engaged with competing information rather than because of a visual obstruction. Indeed, eye-tracking experiments have demonstrated that this phenomenon is robust even when the unnoticed items are directly fixated on, and it occurs even when the stimulus in question is completely apparent to anyone whose attention is not already preoccupied. This entry provides a brief overview of the nature of inattentional blindness, its relationship to related phenomena, and factors that contribute to the likelihood of it occurring.

The Nature of Inattentional Blindness

Inattentional blindness belongs to a family of phenomena that illustrate the central role of attention in visual awareness. Notably, the characteristics that differentiate inattentional blindness from these related phenomena are the same characteristics that make it especially surprising and striking. These characteristics include the degree to which inattentional blindness can be sustained over long periods of time and the degree to which it stems from perceiver-controlled allocation of attention instead of from fundamental limitations of the attention system.

In one of the best known demonstrations of inattentional blindness, participants watched a video of two teams of people, one team wearing black and one team wearing white, interweaving and passing basketballs among themselves. Participants were instructed to count the number of passes made by one of the two teams; partway through the video, a person in a gorilla costume walked through the game, stopped in the middle, pounded her chest at the camera, and walked away, having remained clearly visible for a full 5 seconds. This unusual and striking event was clearly seen by anyone not engaged in the pass-counting task, but fewer than
half of the participants who performed the counting task were aware that it had occurred at all; people’s attention had been so preoccupied with the primary counting task that they simply did not attend to the gorilla and therefore failed to notice it. The fact that the gorilla had been visible for at least 5 seconds and was seen by people not engaged in the counting task underscores the fact that this effect did not stem from fundamental limitations of attention. Indeed, the “gorilla video” and similar work on sustained inattentional blindness have shown that such effects are particularly striking because they can emerge even when stimuli are clearly visible for extended periods of time.

**Perception Without Attention: Inattentional Blindness Versus Visual Search**

Although the line of research that inspired the “gorilla video” stretches back at least to the 1970s, when Ulric Neisser and his colleagues demonstrated similar failures of visual awareness, the term inattentional blindness was coined in the 1990s by Arien Mack and Irvin Rock, who endeavored to discover what aspects of a visual scene are perceived without attention. To provide some background: The notion that some types of features can be perceived without attention had been supported by earlier research employing what is known as a “visual search” task, in which participants search for a predefined target ensconced amid varying set sizes of non-targets. Under normal circumstances, the time it takes to find the target increases as the number of non-targets in the display increases; this result suggests participants needed to attentionally scrutinize each display item to determine whether it was the target. However, the time it took to find certain types of simple features, such as a unique color or orientation, appeared hardly to increase as the number of non-targets increased, suggesting that such features did not require attention to reach awareness. However, Mack and Rock pointed out that visual search experiments cannot eliminate a possible role of attention, as the very act of searching for a target inherently involves attention.

To better assess perception without attention, they developed a task in which participants were retrospectively probed about their awareness of a stimulus that they did not expect. In this task, a cross typically appeared for 200 milliseconds on each trial, and participants’ task was to indicate on each trial whether the horizontal or vertical extent of the cross was longer. For the first several trials, nothing but the cross appeared, but on a critical trial, an additional item appeared in one of the cross’s quadrants. Participants were then probed as to whether they had noticed anything besides the cross on that trial. The results were surprising: Across a number of experiments, about 25% of participants failed to notice the unexpected item—that is, they were inattentionally blind to it—even when it contained features such as a unique color or motion signal, both of which are properties that visual search experiments had previously suggested could be perceived without attention. Interestingly, additional work has suggested that some aspects of perception, such as the grouping of background elements in a display in such a way as to affect perceptual judgments, appear to remain intact even when people cannot report such groupings.

In the context of a field dominated by tasks involving active search for targets and/or the use of “implicit” (i.e., non-self-report) measures such as response time, work on inattentional blindness introduced at least two novel and important factors. First, rather than using implicit measures to make inferences about attention and perception, inattentional blindness experiments probe directly whether or not a stimulus has been consciously perceived. Second, in inattentional blindness experiments, the critical stimulus is unexpected and therefore not actively sought. Notably, the first of these aspects—the necessarily retrospective nature of the probe—has inspired debates about whether inattentional blindness actually reflects a form of amnesia, in which people see the unexpected object but immediately forget it.

**Factors Influencing Inattentional Blindness**

Of critical importance to demonstrations of inattentional blindness to date is the unexpected nature of the critical stimulus. Once people know to expect it, it is difficult to elicit the effect, as the stimulus will no longer be fully unattended. Because of this, inattentional blindness experiments tend to provide one data point per participant; asking participants whether or not they had seen something the first time around clues them in to its probable presence on subsequent trials.
**Stimulus Properties**

Despite the intuitive notion that stimulus properties, such as luminance and color, should greatly contribute to the likelihood of being seen, and despite findings that some such features are rapidly detected during visual search, inattentional blindness research suggests that most features fail to capture awareness automatically (although they can increase noticing rates at least somewhat). Notably, studies have found that stimuli with high degrees of meaning, such as a happy face among healthy participants or a sad face among socially anxious participants, tend to be noticed more readily than nonmeaningful stimuli.

**Attentional Set**

The manner in which people have “tuned” their attention to prioritize certain features over others—in other words, people’s attentional sets—appears to play a very large role. For example, in a computerized series of tasks, participants saw four white and four black shapes move around a display, frequently bouncing off the display’s edges. Half of the participants counted the bounces made by the black shapes (thereby prioritizing the color black) and half counted the bounces made by the white shapes (thereby prioritizing the color white). For the first two trials, nothing unexpected happened, but on the third, critical trial, an unexpected cross (a unique shape in the display) entered from the right side of the display and traveled in a linear path to exit from the left side of the display, passing over a central fixation point and remaining visible for about 5 seconds. The critical manipulation involved the color of the cross, which was either white, light gray, dark gray, or black, so that its color matched that of either the target or the non-target items or fell somewhere in the middle. The results revealed a linear trend, such that the more similar the unexpected cross was to the targets, and the less similar it was to the non-targets, the more likely it was to be noticed (rates of noticing ranged from 0% to 94%, based on attentional set alone). Follow-up experiments demonstrated that attentional set similarly modulated noticing rates regardless of whether participants tuned their attention for shape (e.g., an unexpected square was more likely to be noticed when participants were counting bounces made by other squares than by circles) or for complex assemblages of features such as faces (e.g., grayscale African American vs. Caucasian faces that were matched for brightness).

Recent evidence also suggests that people can tune their attention for abstract meaning in such a way that it influences awareness. Participants viewed a 1-second array of either four words or four small line drawings and then wrote down the identities of the two words or pictures from a given category (e.g., animals or furniture) while ignoring items from a second category. On a critical trial, an unexpected picture or word representative of either the attended or the unattended category appeared at fixation. Participants were more likely to see this unexpected item when it belonged to the same semantic category as the attended, rather than the unattended, category. Importantly, for those writing down the identities of pictures, it was an unexpected word that appeared, and vice versa, so visual similarity was not likely to be a factor.

**Task Difficulty**

The difficulty of the primary task is another factor that influences inattentional blindness rates. For example, when people attempt to perform a difficult perceptual judgment, they are more likely to experience inattentional blindness than when they attempt to perform an easy one. Heavy cognitive load, too, appears capable of inducing inattentional blindness. In one study, an unexpected item appeared as the only item in a display while people tried to hold in mind a number of previously displayed items. Despite the fact that the unexpected item was the only item visible at that moment, it was less likely to be noticed among people who held several items in mind than among those who held only one.

**Related Phenomena**

Inattentional blindness is often discussed in the same context of—and sometimes as if it were interchangeable with—change blindness, which refers to people’s poor ability to notice even large changes that occur in a scene across immediately successive glances. Even though both inattentional blindness and change blindness speak to how much we tend to overestimate our awareness of our surroundings, they are quite different phenomena. Whereas inattentional blindness refers to failures to notice objects and events in the first place, change blindness refers
to failures to notice changes to things despite our already having registered their appearance.

Less frequently confused with inattentional blindness, but more closely related to it in some respects, is the *attentional blink*. The attentional blink occurs when people monitor a rapid stream of stimuli (e.g., with stimuli presented for a 10th of a second or less) in order to report two targets. For example, participants might be required to report the two digits appearing within a stream of letters. When the two targets appear close together in the stream (e.g., separated by only a couple of intervening items), people often can report the first target but not the second one; however, when the two targets appear far apart in the stream, participants generally can report both. The similarity to inattentional blindness lies in the demonstration that failures to notice a stimulus occur in the absence of attention (in this case, when attention is preoccupied with the first target). However, in contrast to inattentional blindness, the phenomenon stems from limits in the temporal resolution of attention and therefore occurs even when people know to expect both targets.

Inattentional blindness also has strong connections with the literature on *attentional capture*, in which implicit measures such as response times and eye movements have been used to suggest what types of features grab attention without a person’s volition. Although the weight of evidence suggests that attention is necessary for conscious perception, there appears to be a dissociation: Stimuli that have been shown to attract attention via implicit measures have been shown not to similarly grab awareness per se. Indeed, there are examples in the literature where implicitly measured shifts of attention have been observed despite participants not being aware of the attention-capturing stimulus. Such findings suggest that although attention may be necessary for conscious perception, the allocation of attention can also occur without leading to conscious perception.

*Steven B. Most*

**See also** Attention and Consciousness; Attentional Blink Effect; Change Blindness; Divided Attention and Memory; Perceptual Consciousness and Attention; Visual Search

**Further Readings**


**INDEXICAL THOUGHT**

This entry will describe some of the main philosophical and psychological puzzles that arise concerning the kinds of thoughts that are expressed using indexical terms such as *I, here, and now*. These will include some puzzles concerning the relation between perspectival and objective modes of thought, the significance of indexical thoughts for the thinker’s actions, the possibility of the same indexical thought being entertained by different thinkers or by the same thinker at different times, and the nature of thoughts about oneself.

An indexical is a linguistic expression whose reference varies in a rule-governed way with the context of utterance. Standard examples include *here, now*, and *I*, which normally refer to the place, time, and speaker of the utterance, respectively. Indexical thoughts are thoughts expressible using indexicals. *Indexicals*, as the word is used here, are to be distinguished from *demonstratives* such as *that*, whose reference also varies with context but not in the same rule-governed manner (note, however, that some writers use *indexical* or *demonstrative* for all context-sensitive terms). There is much debate about communication using indexicals (e.g., concerning messages saying “I am not here now”) but because indexical thoughts need not actually be expressed, or might be expressed only in internal monologue, we can largely set these problems of communication aside.

Indexical thoughts raise many puzzles. Some concern the fact that to represent the world indexically is to represent it from one’s own perspective or point of view. This is often described in terms...
of egocentricity. To think of a place as nearby and to the left, for example, is to think of it as an egocentric location in an egocentric frame of reference determined relative to the thinking subject. Indexical thoughts are subjective in that they represent the world from the subject's point of view. Perception is similarly perspectival. Yet we can describe many states of affairs without using indexicals, and scientific descriptions of the world should normally avoid indexicals. Much of our conception of reality thus seems to require that we possess map-like or allocentric representations of the world, that is, representations from no particular point of view.

One qualification should be mentioned: *Now* is an indexical because its reference varies according to the time of utterance. But according to some theories in metaphysics (A-theories), *past*, *present*, and *future* are real, objective properties of times, whereas according to other theories (B-theories), they are relations analogous to *left*, *up*, or *near*. Hence, if the A-theory is correct, temporal indexical thought represents the world in an objective, non-egocentric way, whereas if the B-theory is correct, temporal indexical thought is egocentric in the same way as spatial indexical thought.

Indexical thoughts have a special significance for action. In John Perry's famous example, when one notices a trail of sugar along the supermarket floor leaking from someone's shopping trolley one thinks "someone is making a mess." But when one realizes that the sugar is in fact leaking from one's own trolley one thinks "I am making a mess," and one acts in a different way. Similarly, believing that the meeting begins at 3 p.m. does not have the same significance for one's actions as believing that the meeting begins in two minutes, *or now*, *or half an hour ago*. Merely knowing that one should head north is of no practical use in itself; one needs to know which way north is in egocentric terms, for example, *to the right*. Thus, as Perry observed, indexicals are essential: One cannot adequately describe an indexical belief-state using only non-indexical vocabulary, and perhaps one cannot act at all unless one represents the world indexically.

Indexical thoughts also raise questions about the individuation of thoughts. In another of Perry's examples, suppose that I believe, correctly, that "a bear is about to attack me." You see what is happening and say, "A bear is about to attack you." In one sense, you and I believe the same thing. Yet there are systematic differences in our behavior; I roll up in a ball and keep still while you run to get help. Because the individuation of beliefs should capture their psychological and behavioral role, this suggests that in fact you and I have different beliefs, associated with the different indexicals. However, there are still problems. Suppose I believe that "Napoleon gave a speech here." I walk away, but keep track of my movements through perception. I now believe "Napoleon gave a speech there," where *there* refers to the place previously referred to as *here*. According to the argument given earlier, this is a new, different belief. But haven't I retained a belief, and retained it in a more psychologically significant way than in the sense in which you and I shared the "same" belief about the bear? This is the problem of cognitive dynamics, raised by David Kaplan. A related problem affects communication. It is sometimes held that communicating involves sharing one's thoughts with one's audience. But when I say *I* (*or here*), you replace it with *you* (*or there*). Does this show that you can never believe what I believe about myself (*or about here*)?

It is sometimes thought that a complete account of first-person thought would constitute a complete account of the "self." Much work has focused on various unusual epistemological features of first-person thought. Among these is the phenomenon of immunity to error through misidentification: There are certain ways of ascribing a property to myself such that, even if I could be mistaken about whether I instantiated the property, this could not be because I was mistaken about *nobo* instantiated the property. Although this phenomenon is arguably shared by some other indexical and demonstrative judgments, a better account of it may nonetheless help illuminate the nature of first-person thought.

*Simon Prosser*

**See also** Concepts and Language; Descriptive Thought; Object-Dependent Thought; Self-Consciousness; Self-Knowledge

**Further Readings**


**INFERENCES IN LANGUAGE COMPREHENSION**

Consider the first clause of Lewis Carroll’s *Alice in Wonderland*: “Alice was beginning to get very tired of sitting by her sister on the bank and of having nothing to do.” This sentence, although apparently simple, conveys innumerable implications. Somewhat trivially, it implies that Alice was sitting by a related, female human being. More meaningfully, we interpret Alice to be bored and perhaps near sleep. Although unstated, we readily judge that Alice is near a river and not atop a financial institution.

Proper comprehension of coherent messages, called *discourses*, requires the understander to draw inferences about many of these unstated ideas. The present focus is on inferences that combine discourse meaning and people’s ordinary world knowledge. Inferences that maintain coherence by bridging the parts of a message will be distinguished from inferences that usefully elaborate discourse meaning but are not comparably necessary. Both types promote people’s mental construction of the characters, time, locations, and causes underlying the situations described in ordinary messages.

**Levels of Language Representation**

Understanders mentally represent language messages at multiple levels. First, messages convey idea units, called *propositions*: One idea unit of the *Alice* sentence is that Alice sat by her sister. Understanders integrate message propositions into a network. Second, people construct mental representations of the situations communicated by a message. Consider the sequence, *The Queen outweighed Alice. Alice outweighed the Hatter*. A postreading judgment that the Queen outweighs the Hatter must refer to a situation model, because the message did not directly compare those two characters.

Understanders also encode the precise, verbatim form of messages but those representations fade much faster than proposition networks and situation models unless they are of special interest or relevance to the understander. Language inferences are viewed as predominantly encoded in situation models although possibly in proposition networks as well.

**Bridging Inferences**

The sequence, *The tooth was pulled painlessly. The dentist was pleased*, is sensible. However, the phrase *the dentist* suggests prior knowledge about a particular dentist; but the first sentence has not mentioned one. The peculiarity of the alternative continuation, *The architect was pleased*, highlights this comprehension puzzle.

Understanding the *tooth* sequence requires an inference that *bridges* the two sentences: The proposition that the tooth was pulled might be inferentially augmented with the role, DENTIST. The suitability of this inference is confirmed by our ordinary knowledge about dental health. Without bridging inferences, message coherence would be disrupted.

The encoding of bridging inferences is reflected by judgments, during and after comprehension, that are frequently as accurate and fast for stimuli that capture bridging inferences as for ones that repeat explicit message ideas. That includes tasks such as (a) judging the truth of *a dentist pulled the tooth*, and (b) naming (reading aloud) a bridging inference concept such as BROKE after reading *The delicate vase fell. Its replacement cost $100.*

**Anaphoric Bridges**

Anaphors are expressions that refer to prior message elements. The pronoun is the prototypical anaphor. Thus, *her in sitting by her sister* in the opening excerpt refers to Alice. Pronouns bridge current and antecedent text ideas, guided by the gender, number, and person of the pronoun. Anaphoric bridging processes accompany comprehension. Accordingly, it takes less time to read (a) *Mary checkmated Ron because she was a grandmaster*, in which *she* has only one possible antecedent, than (b) *Mary checkmated Helen because she was a grandmaster.*
Inferences in Language Comprehension

Situational Bridges

Situation model representations of discourse reflect dimensions such as cause, time, space, characters, and characters’ goals. Encoding these relations often demands subtle bridging inferences by the understander. One index of these bridging processes is greater reading time when a phrase is inferentially inconsistent than consistent with prior situational information. Thus, spatial bridging processes are exposed by greater reading time for *The instructor came in* after having read that Jane was waiting *outside* the gym rather than *inside*. Likewise, the bridging of time relations is indicated by longer reading time for *Claudia was already waiting for Markus* after reading that Claudia’s train arrived *after* (rather than *before*) Markus’s.

Readers’ bridging along situational dimensions is also revealed by faster judgments about inferential than noninferential concept words. Thus, recognizing that *gift* has been previously mentioned takes less time when a character’s *goal* to buy a gift had previously failed rather than succeeded. This suggests that the reader is inferentially monitoring the uncompleted goal and that inferences significantly depend on memory for prior discourse ideas. Furthermore, sentence reading time increases systematically with the number of required bridging inferences, summed across situational dimensions. This indicates that comprehension typically yields a complex, multidimensional model. Focusing on one dimension may enhance its impact on comprehension but generally does not prevent the encoding of others.

Elaborative Inferences

On reading the isolated sentence, *The tooth was pulled painlessly*, inferring that a dentist participated is a sensible elaboration, but it does not bridge different message parts. Early research surprisingly indicated that even highly probable elaborative inferences are not reliably encoded—it took longer to judge sentences and words representing elaborative inferences than corresponding bridging inferences and explicit statements. This is likely because the sheer volume of sensible elaborations, as suggested by the *Alice* excerpt, would overwhelm the cognitive resources of the understander.

However, elaborative inferences that are strongly constrained by their message appear to be encoded during comprehension. For example, *they traveled overnight from Vancouver to Beijing* supports the elaborated role, AIRPLANE. People infer the highly constrained properties of discourse concepts: *The giraffe’s height amazed the boy* implies TALL. *Pete positioned the thumbtack on the chair* yields the elaborative generalization, PRANK.

Predictive Inferences

Elaborative inferences about understanders’ predictions about the outcome of discourse events have been extensively scrutinized. For example, *The vase fell to the concrete* suggests the prediction that the vase has broken. A complex and informative picture emerges from the research. Immediately after reading a sentence, people’s judgment times about predictive-inference words (*broke*) are less than after control sentences, signifying inference activation. However, unless the next sentence corroborates the inference or requires a bridging inference about it, its activation diminishes and it is unlikely to be permanently encoded.

The initial activation of elaborative inferences is regulated by the current sentence plus its interactions with prior discourse context, such as suggestions of an alternative outcome or the pertinent trait of a character. The reader’s intention to predict discourse outcomes accelerates the activation of elaborative inferences. In general, activation of elaborative inferences likely provides the building blocks for encoding robust, enduring inferences.

Theories and Factors of Inference

A Process Model

A unified inference analysis has been derived from the construction-integration (CI) model of the renowned investigator, Walter Kintsch. From the CI perspective, understanding a phrase first involves the construction of a network of its explicit propositions, plus its implications, close associations, and generalizations. A subsequent, integration processing phase enhances the activation of highly interconnected network ideas but deactivates sparsely connected ones.

The CI inference analysis proposes that sentence implications are represented (activated) predominantly in the situation model but also in the sentence’s proposition network. Subsequent phrases that must
be bridged to that sentence connect with its implications and serve to sustain them. Lacking such bridging connections, many elaborative inferences may become deactivated and abandoned. The exception to this is when other discourse constraints, such as the aforementioned roles, properties, and generalizations, consolidate tentative elaborative inferences.

**Factors Influencing Language Inferences**

Many factors collaboratively guide people's language inferences. Messages differ in the importance and interestingness of their implications and the distance between inferentially related ideas. People vary in their cognitive resources, basic language skills, and previous knowledge about a message. People have comprehension goals such as learning, memorizing, or being entertained. Text genres such as the narrative, exposition, and description each bear unique linguistic devices and structures. These variables have been shown to regulate language inference in complex, interactive ways.

_Murray Singer_

**See also** Conversation and Dialogue; Discourse Processing, Models of; Working Memory in Language Processing

**Further Readings**


**INHIBITION OF RETURN**

Inhibition of return (IOR) is an effect whereby people are slower to respond to a stimulus presented at a recently stimulated or inspected location compared to a stimulus presented at a previous unstimulated or uninspected location. The IOR effect is often studied using a cue-target paradigm (see Figure 1), made popular by Michael I. Posner and colleagues. In the cue-target paradigm, participants are asked to fixate a central cross, flanked by a square on each side. The first stimulus, referred to as the cue, is a change in luminance of one of the two squares. A short time later, a second stimulus, referred to as the target, is presented inside one of the two squares. Both the cue and the target are equally likely to appear at both locations. Participants are not required to respond to the cue, which is not informative about either target location or identity. Participants can be asked to perform different tasks, such as detecting the target (pressing a key wherever the target is presented), localizing its position (reporting the location of the target), or discriminating some of its features (such as its color or shape). Both speed of response and accuracy are measured as dependent variables.

IOR usually appears at long cue-target intervals. At shorter intervals (~100–500 ms), a brief period of facilitation is observed, during which time people are faster to respond to a target preceded by a cue at the same location. The canonical interpretation of this pattern of results is that attention is briefly captured by the cue and that, once attention has left the cued location, it is slower to return, hence the name *inhibition of return*. This entry will focus on describing some essential characteristics of the IOR effect, its possible purposes, and the mechanisms underlying its functioning. Finally, the entry will explore the consequences of IOR in visual processing as well as its neural substrates.

**Characteristics of the IOR Effect**

Several characteristics of the IOR effect have been recognized by using variants of the standard cue-target paradigm. When the task involves either target detection or target localization, the IOR effect is observable at cue-target time intervals of around 300 milliseconds. When the task involves target identification, IOR appears at longer cue-target intervals of 500 milliseconds or longer. The IOR effect is relatively long lasting, and only begins to diminish when the cue-target interval is approximately 3 seconds or longer. It has been established that IOR can be environmentally coded, such that the IOR is observed...
Inhibition of Return

at the environmental location where a cue was presented rather than at the location stimulated on the retina. The IOR effect also appears to operate on both moving and stationary objects and can co-occur at up to five locations when multiple cues are presented sequentially.

Whereas the IOR effect is typically observed using peripheral cues and targets while having participants fixate a central cross, IOR is also observed when eye movements are involved. For example, when participants are required to move their eyes to the cue and then return them back to the center before the target appears, both manual responses and eye movements to a target are slower if the target is presented at the previously inspected location. IOR is also measured for consecutive responses in a so-called target-target paradigm. When participants have to respond to successive targets presented in the periphery, their response times are slower to detect targets presented at the same location than the previous one. The IOR effect can be cross-modal—people are slower to respond to a target at a previously cued location, even when the cues and targets are presented in different sensory modalities.

Why IOR? Possible Purposes of the Effect

Attention seems to be automatically and reflexively captured by unexpected peripheral stimuli. However, it would be disadvantageous for attention to be repetitively drawn to the same location, especially if this location does not contain relevant information for our goals. In his seminal work, Posner suggested that the purpose of the IOR effect is to bias attention toward novel locations. Taking this suggestion a step further, Raymond Klein proposed that the IOR effect serves to facilitate foraging by preventing reinspection of recently inspected locations while searching the environment. In support of this notion, it has been shown that when people are free to search a dense display for a target stimulus, they are slower to detect an unexpected dot in the search array when it is presented at a recently inspected location compared to when it is presented at a new, previously un inspected location.

These related hypotheses assume that IOR plays a role in selective attention. Selective attention refers to the notion that we have limited resources for processing the information in our environment, and so we must select which information will receive further processing. In vision, it has been hypothesized that information is represented in a salience map. The most salient or noticeable information is given priority for selective processing. The salience of an item is based on both the goals of the observer and the item's physical features; the item with the highest level of salience captures attention. Following inspection of the item, its salience must be reduced in order to prevent continuous fixation. Some researchers have suggested that the role of IOR is to reduce the salience of recently inspected items so that new items may be inspected.

How Does IOR Happen?

Underlying Mechanisms

The mechanism or mechanisms underlying IOR are still an unresolved question. Here are some proposals:

Inhibition of Attentional Orienting

This is the hypothesis after which the IOR effect was named. It postulates that IOR is due to the inhibition of the return of attention to a previously attended location. This hypothesis assumes that attention needs to be withdrawn from the cued position for IOR to be generated. Although this hypothesis is implicit in the understanding of the IOR effect for most researchers in the field, the recent finding that IOR is observed even when attention is maintained at the cued location belies this idea.
Inhibition of Return

Oculomotor Activation

In the classical procedures, the eyes remained fixated at the center while the cue and the target were presented. It was proposed that the peripheral cue automatically activated the oculomotor system, and IOR was caused by its inhibition in order to maintain central fixation. It was soon discovered that inhibition of the oculomotor system was not necessary to observe IOR, given that IOR was found even when the eyes moved to the cue, and therefore, no such inhibition was involved. Furthermore, IOR was observed even when participants were directly looking at the target, and therefore the oculomotor system was not inhibited. Currently, this hypothesis postulates that the activation of the oculomotor system (but not its inhibition) is necessary to observe IOR.

Habituation of the Orienting Response

The orienting response refers to the shift of appropriate sensory receptors toward an unexpected stimulus in the environment that may have some importance. Repetition of the stimulus without consequence results in a decrement of the orienting response called habituation. Habituation is a form of learning in which an organism determines that the presentation of a stimulus is not paired with a consequence and therefore learns to ignore it. As with IOR, habituation of the orienting response is thought to promote novelty by preventing reorienting to repetitive stimuli that are without consequence. It has been suggested that some forms of IOR are actually examples of habituation of the orienting response.

Loss of Novelty

This hypothesis postulates that our attention system is biased to detect new events. When a location has already been inspected, it loses its novelty, resulting in a cost in detecting new objects at this location versus uninspected locations. The major difference between this explanation and the habituation explanation is that the act of learning is removed from the equation. As such, this explanation predicts that if cues were made relevant by having them indicate the location of the target on more than 50% of trials, IOR would remain unchanged because a target presented at the uncued location would still be novel in comparison to a target presented at the same location as the cue. The habituation explanation suggests that making the cues predictive of the target location would diminish or eliminate IOR, because the cues are no longer irrelevant to the task.

Consequences of the IOR Effect

Whatever its causes, IOR can affect different stages of information processing, including the sensory-based perceptual stages, the response-generation or motoric stages, and the higher decision-level stages. When a standard peripheral cue-target paradigm is employed and eye movements are prevented, the perception of new objects at the cue location is impaired. Event-related potentials have also revealed some of the neural consequences of the IOR effect. Event-related potentials are electrical brain responses measured using noninvasive electrodes placed at various points along the scalp to infer activation of groups of neurons. The activation is measured and averaged over many trials to obtain a clear signal based on specific events presented during the experiment. In a peripheral cue-target paradigm, the event-related potentials associated with activations of the extrastriate visual areas are diminished at previously inspected versus uninspected locations, indicating that IOR impairs the perceptual detection of new events. Event-related potentials have also been used to demonstrate that IOR affects pre-motor processes associated with response preparation, suggesting that IOR can affect participants’ readiness to respond. IOR has also been shown to affect the criterion that participants adopt for responding, manifested as participants being more reluctant to respond to previously inspected compared to uninspected locations. However, IOR does not always produce an effect on all of the stages of processing, and the effect produced largely depends on the task at hand.

Neural Correlates of the IOR Effect

Subcortical Correlates

IOR has been strongly linked with subcortical brain structures. The most important evidence in favor of the relevance of subcortical structures in IOR, and particularly of the superior colliculus, is the case of patients who suffer from a degenerative disease, progressive supranuclear palsy, affecting this structure. In the initial stages of the disease, vertical eye movements are impaired, while horizontal eye
movements are spared. IOR is selectively impaired in these patients in the vertical but not in the horizontal axis, thus supporting the important role of the superior colliculus in the generation of IOR. A further piece of evidence in relation to the importance of subcortical structures for the effect is that IOR is observed in newborns, for whom cortical structures are not fully developed.

**Cortical Correlates**

It has been shown that cortical pathways also play an important role in the generation of IOR. When participants are presented with cues that have been shown to bypass the superior colliculi, such as low wavelength purple stimuli, IOR is not observed when participants have to respond by moving the eyes. However, using these same stimuli to which the superior colliculi are blind, IOR is observed when participants have to manually respond to the target. This indicates that the underlying processing circuitry of IOR is dependent on the type of motor response used for its evaluation, so that when responses are not oculomotor but manual, structures other than the superior colliculi might be implicated in its generation.

One of the cortical structures that seems to play an important role in IOR is the frontal eye fields. This structure, which is located in the frontal cortex, is related to the voluntary generation of eye movements and is associated with the orienting of attention. The parietal cortex is also important for the generation of IOR. Regions of the parietal cortex receive converging information that is integrated in saliency maps, which determine the location to which attention will be oriented at any given time. The already inspected locations will be impaired in these saliency maps, producing a cost in detecting targets at previously inspected locations. The parietal cortex in the right hemisphere of the brain seems to be especially relevant in the formation of these saliency maps. When it is damaged or disconnected, patients develop a syndrome known as spatial neglect, in which IOR does not occur for targets presented to the right hemisphere.

**Further Readings**


**INNATENESS AND PARAMETER SETTING**

There have been major advances in linguistic theory in the past 50 years, and some of them have led to corresponding advances in understanding how children acquire language. Advances on both fronts resulted from the shift from rule-based theories of grammar in the 1960s and early 1970s, to the principles and parameters theory of the 1980s and 1990s, and its descendant, the minimalist program. According to the principles and parameters theory, children are not expected to accrue individual rules governing the local language, as had been supposed using earlier theories of syntax.

The principles and parameters theory supposes that children are biologically fitted, as part of the human genome, with a universal grammar. Universal grammar contains the core principles of language,
that is, principles that are manifested in all human languages. These principles established the boundary conditions for grammar formation. Children were not expected to deviate from these principles in the course of language development. In addition, universal grammar spells out particular ways in which human languages can vary. These points of variation are called parameters.

Once parameters were entered into the theory of universal grammar, many cross-linguistic differences that had previously been assumed to be learned were reconceived as innately specified properties of human languages. So parameter setting is largely a matter of genetic specification, along with a minimal, but crucial, contribution from the environment. Because both the principles and parameters of universal grammar are innately specified, the principles and parameters theory is a “nativist” approach to language acquisition, to be contrasted with “nurture” approaches, which view language acquisition as largely the product of children’s experience. The principles and parameters theory enabled researchers in child language to formulate and evaluate far-reaching predictions about the course of language acquisition, including predictions that were not consistent with experience-based accounts. This entry ends with a sample of the kinds of insights into child language acquisition that have been achieved within the principles and parameters theory.

The introduction of parameters was designed to make language learning easier. This advanced the theory of universal grammar along its stated goal of explaining children’s rapid mastery of human language. Parameters were seen to reduce the role of experience in children’s acquisition of human languages. Parameter setting was seen to initiate radical changes in child language. Instead of piecemeal acquisition of specific constructions, as advocated by experience-based accounts of language development, acquisition was seen to involve the mastery of clusters of linguistic phenomena, all in one fell swoop. The metaphor that was often coupled with descriptions of parameter setting was that of a switch, as in a circuit box. The learner set the switch one way or the other in response to “triggering” experience. If the switch was set one way, then the child’s grammar assumed one format; if the switch was set the other way, the child’s grammar assumed a different format.

Although experience was a prerequisite to setting parameters, it is important to appreciate that parameter setting did not have the character of learning. Instead, parameters were assumed to be set reflexively. Reflexive responses by a species to particular inputs are characteristic of genetically determined acquisition in any domain. Similarly, the expectation was that parameter values were triggered or fixed by experience, as in one-trial learning. This accounted, in part, for the rapidity of children’s acquisition of language, as compared with other cognitive skills such as the ability to learn to count, or to draw, or to play a musical instrument.

Reducing the role of experience by invoking parameters raised new questions, including whether children should be expected to adopt an initial, default parameter value. For some parameters, the answer was affirmative. The initial value for these parameters was determined by a learning mechanism, the subset principle, originally proposed by Robert Berwick. The subset principle ordered parameter values according to the number of linguistic structures (and corresponding meanings) that could be assigned to a linguistic expression. The subset principle became operative when the class of languages that adopted the alternative setting of P (call these P1 languages), generated fewer structures for a given type of expression than the class of languages that adopted the alternative setting of P (call these P2 languages). In such cases, the subset principle compels children learning P1 languages and children learning P2 languages to hypothesize initially that the local language is a P1 language rather than a P2 language. Children learning P2 languages are therefore expected to speak a fragment of a “foreign” language for a while.

To recap, experience is needed to set parameters. However, to the extent that language variation is explained by innately specified parameters, the role of experience is minimized, as compared to its role in experience-based models of language development. By reducing the imprint of experience on language learning, parameters were another vehicle (in addition to core linguistic principles) for guaranteeing children’s convergence on “correct” grammars of the local language, that is, grammars that are sufficiently similar to those of other speakers, so as to permit effortless and seamless communication.

It had been assumed until quite recently that the language apparatus used in setting parameters was specific to language. This assumption has been called into question, however, in the theoretical descendant of the principles and parameters framework, the
minimalist program. The possibility that parameter setting might involve cognitive principles outside the language faculty is being taken seriously within the minimalist program. For example, empirical findings about children's abilities to mine data, as well as considerations of computational efficiency, have been instrumental in leading researchers to reexamine the nature of parameters.

**Parameters in Child Language**

This section describes two parameters. The first parameter illustrates how changing the setting of a parameter can have wide repercussions in children's grammars. The second parameter invokes the subset principle. Both parameters reveal differences between child and adult language, phenomena that resist explanation on experience-based accounts of language development and comport better with the nativist approach.

### The Negation Parameter

Young children's negative statements bear little resemblance to their adult counterparts. Children's negative statements often contain *not* (e.g., *I not hurt him*), whereas the corresponding statements by adults (*I didn't hurt him*) contain auxiliary verbs (e.g., *do*) with an appended clitic form of negation *n't*. English-speaking adults use *not* for sentential negation only about 10% of the time, whereas some English-speaking children do not begin using the clitic form of negation *n't* until after age 3. The negation parameter was postulated to explain this mismatch between caretaker input and children's output. The present version of the negation parameter is based on work by Hedde Zeijlstra and by Rosalind Thornton and Graciela Tesan.

Across languages, negation can either be an adverb or the head of the negation phrase, labeled NegP (a noun is the head of a noun phrase, NP; a verb is the head of a verb phrase, VP; and so on). The initial value of the negation parameter analyzes negation as an adverb. Children do not project a phrasal category such as NegP unless they encounter evidence that this category is required in the local language. If the negation phrase NegP is initiated by relevant input, children posit that negation is the head of that phrase and not an adverb after all.

There is abundant evidence that standard English requires a negation phrase, because the clitic form of negation *n't* in negative auxiliaries (e.g., *can't, don't*) resides in the head of NegP. Why, then, do children produce negative statements with *not* until they are 3 years old, rather than adult-like statements with *can't and don't*? The answer is that the internal morphology of negative auxiliary verbs is not transparent to children younger than 3 years. This delays children's convergence on the adult grammar and explains why there is a transitional period during which they analyze *can't* and *don't* as adverbs. These words form unanalyzed chunks for children, which function just like the adverb *not*.

The negative auxiliary verb *doesn't* resists this analysis. *Doesn't* has third person agreement internal to the word (*I do, you do, he does*). This fact encourages children to decompose the negative auxiliary verb *doesn't* into three heads: *do + s + n't*. The appearance of *doesn't* may, therefore, be critical evidence for children to add a negation phrase NegP to their grammars. Adding NegP, in turn, causes children to switch to the alternative value of the negation parameter, according to which negation is a head rather than an adverb.

During the initial period, when negation is an adverb, child and adult language differ in several ways. One difference was mentioned earlier: Children's early negative utterances contain *not* where adults use negative auxiliaries and the clitic form of negation *n't*. A second difference involves the tense/agreement marker. Where adults would say *It doesn't fit*, children younger than 3 years old permit *It not fits*. Here, children are associating the tense/agreement marker with the main verb.

A third difference is that children avoid negative questions like *What don't you like?* In an elicitation production study, the most frequent structure produced by the 4-year-old children contained auxiliary doubling, for example, *Do you don't like cheese?* and *What do you don't like?* In these questions, the auxiliary verb *do* precedes the subject noun phrase *you*, and the negative auxiliary *don't* follows it. These non-adult negative questions are expected if negation is analyzed as an adverb, and if *don't* is also an adverb, and not morphologically decomposed.

A fourth difference concerns children's use of negative auxiliary verbs in tag questions. In this construction, a statement is followed by a “tag” that turns it into a question. For example, *You like cheese, don't you?* bears the tag *don't you*, which is a reduced yes/no question, with the auxiliary preceding the subject NP. Because young children avoid...
subject-auxiliary inversion, it is not surprising that children's early productions lack tag questions.

**Resetting the Negation Parameter**

Recall that the trigger for English-speaking children to reset the negation parameter is the negative auxiliary verb *doesn't*. This provides children with evidence that the local language projects a negation phrase and, consequently, that negation is a head. Adding a negation phrase initiates parameter resetting. Once the parameter has been reset, all of the earmarks of the adverbial setting gradually disappear and are replaced by adult-like uses of negation. As soon as children begin producing *doesn't*, there is an abrupt transition from the negative adverb *not* to the clitic form of negation *n't*. Other changes to children's grammars take longer, however. Until they are 4 years old, children continue to produce non-adult negative questions and continue to avoid producing tag questions. Perhaps the developmental lag in these structures is due to children's inherent conservatism in trying out new structures.

Once children reset the parameter, sentences exhibiting negative concord are, in principle, licensed by children's grammars. Negative concord refers to sentences in which two negative words combine to form a single negative message, such as *Nobody don't like me* (meaning *Nobody likes me*). Children acquiring standard English do not receive input for such sentences. Nevertheless, Ursula Bellugi reports that two of the children whose transcripts she studied, Adam and Sarah, spontaneously produced sentence structures with negative concord when they were 4 years old (after the auxiliary system had become productive). For example, Adam produced examples like *He never won't scare me* and *I'm not scared of nothing*; Sarah produced examples like *Because nobody didn't broke it* and *I don't got no paper today*. The corresponding adult input would include a single form of negation (e.g., *I'm not afraid of anything. Nobody broke it*). Why would children produce sentences with negative concord despite the lack of evidence for this structure in the parental input? Apparently, once children have established that negation is a head, structures with negative concord follow naturally. This supports the claim that negative concord is a default form of negation in child language, which explains why negative concord is a prototypical feature of creole languages.

**The Disjunction Parameter and the Subset Principle**

Before introducing the disjunction parameter, it will be useful to explain the concept of *scope*, which can be illustrated with a simple example. At Logan Airport in Boston, there is a sign with an ambiguous message due to the different scope relations between the words *every* and *not*. The sign reads *Every airplane does not carry pets*. On one reading, *every* takes scope over *not*. This reading can be paraphrased as *None of the airplanes carry pets*, meaning that *every* airplane is such that it does *not* carry pets. On the second reading, *not* takes scope over *every*. This reading can be paraphrased as *Not every airplane carries pets*.

Across languages, words for negation (*not, no*) assume different scope relations when they combine with words for disjunction (*or*). This cross-linguistic variation is attributed to the disjunction parameter, first proposed by Takuya Goro. In keeping with the subset principle, children have been found to initially favor the value of the disjunction parameter that makes sentences true in the narrowest range of circumstances. This ensures that children will have access to positive evidence if the local language favors the alternative scope possibility, one that make sentences true in a broader range of circumstances. This means that children acquiring some languages are expected to initially assign a value to the disjunction parameter that differs from the value of the parameter that is assigned by adult speakers. Consequently, children and adults should interpret sentences that are governed by the disjunction parameter in different ways.

Suppose, after lunch, your friend informs you: *Ted did not order pasta or sushi*. As an English speaker, you understand your friend's statement to mean two things: *Ted did not order pasta* (~P) *and* *Ted did not order sushi* (~S). This interpretation of disjunction in the scope of negation conforms to de Morgan's laws of propositional logic. It is called the conjunctive entailment of disjunction in negative statements. In logic, the formula corresponding to *Ted did not order pasta or sushi* is ~(P or S), which logically entails the negation of both disjuncts (~P & ~S).
1. Ted didn’t order sushi or pasta.

In English, negation takes scope over the disjunction word (or) in negative statements like (1). This is one value of the disjunction parameter.

In contrast to English, adult speakers of Mandarin do not generate a conjunctive entailment when the Mandarin word for disjunction (huozhe) appears in simple negative sentences. Mandarin adopts the alternative value of the disjunction parameter, as compared to English. The translation of the English sentence (1) into Mandarin is (2).

2. (Wo cai) Ted meiyou dian shousi huozhe yidalimianshi.

(I guess) Ted not order sushi or pasta.

“It’s sushi or pasta that Ted did not order.”

In Mandarin, disjunction (huozhe) takes scope over negation (meiyou) in negative statements like (2). Mandarin-speaking adults accept this sentence if Ted ordered pasta but not sushi, and they accept it if Ted ordered sushi but not pasta. Lastly, adults accept the same sentence if Ted ordered neither pasta nor sushi. Recall that this last circumstance corresponds to the unique reading of the English sentence (1). So the English sentence (1) is true in just one circumstance, ~P & ~S, whereas its Mandarin counterpart (2) is true in this circumstance, but it is also true in other circumstances as well, where ~P or ~S. So, the circumstances in which the English sentence (1) is true, ~P & ~S, comprises a subset of the circumstances in which the Mandarin sentence (2) is true. Therefore, the subset principle should be operative when children are setting the disjunction parameter.

The subset principle dictates that, whenever parameter values are in a subset/superset relation, children must initially select the subset value, to avoid the proliferation of interpretations beyond those of the target language. Thus, the subset principle anticipates that children’s initial (default) setting will be the subset (“both not”) reading. If so, Mandarin-speaking children should initially interpret (2) in the same way as English-speaking children and adults interpret (1), despite the absence of evidence for this interpretation in the adult input. Researchers in child language have investigated this prediction, and the findings were exactly as anticipated. Whereas Mandarin-speaking adults generally accepted sentences like (1) in circumstances corresponding to the “not both” reading, Mandarin-speaking children rejected them in these circumstances the vast majority of the time. For children, the Mandarin word for disjunction (huozhe) received a “both not” interpretation, just as or does in English. For adults, by contrast, disjunction received a “not both” interpretation.

The observed differences in interpretation by children and adults are difficult to explain on an experience-based account of language development. Clearly, Mandarin-speaking children were not basing their interpretation on the parental input. Children effectively ignore some of the primary linguistic data, at least for a while, as they follow the natural seams of human languages. At any given time, children are speaking a possible human language, but not necessarily the language that is being spoken around them. Such findings are evidence for the innateness of parameters, as anticipated by the theory of universal grammar.

Stephen Crain and Rosalind Thornton

See also Language Development; Planning in Language Production; Production of Language

Further Readings


**INNER SPEECH**

When asked to reflect on their subjective experience, people often report that it involves a voice or voices speaking internally. Inner speech (also called verbal thinking, covert self-talk, internal monologue, and internal dialogue) has been proposed to develop in childhood through the gradual internalization of dialogues with other individuals, resulting in an internalized “conversation” with the self. Inner speech has been argued to be theoretically important for the development of verbal mediation of cognitive processes in childhood, self-awareness, and psychiatric symptoms such as auditory-verbal hallucinations. Future research priorities include further specification of the mechanisms of internalization, methodological improvements in the study of covert and partially internalized speech, and progress in the measurement of inner speech through experience sampling, dual-task paradigms, and neurophysiological methods.

**Vygotsky’s Theory of Inner Speech**

Although philosophical ideas about inner speech date back as far as Plato, the fullest theory of its development was put forward by Lev Semonovich Vygotsky. Vygotsky proposed that inner speech is the developmental outcome of a process of internalization, through which social speech with others is transformed into an internal dialogue. Vygotsky saw support for his theory in children's overt self-directed speech (private speech) during cognitive tasks, viewing it as a transitional stage in the transformation of interpersonal dialogues into intrapersonal ones. Private speech and inner speech thus have roles in the self-regulation of cognition and behavior, with children gradually taking on greater strategic responsibility for activities previously requiring the input of an expert other.

The transition from social to private to inner speech is accompanied by significant syntactic and semantic transformations, notably the syntactic abbreviation that results in inner speech having a “note-form” quality compared with external speech. Semantic transformations identified by Vygotsky include the predominance of sense over meaning (whereby personal, private meanings achieve a greater prominence than conventional, public ones), the process of agglutination (the development of hybrid words signifying complex concepts), and the infusion of sense (whereby specific elements of inner language become infused with more semantic associations than are present in their conventional meanings). It has been suggested that inner speech should take two distinct forms: expanded inner speech, in which internal dialogue retains many of the acoustic properties and turn-taking qualities of external dialogue, and condensed inner speech, in which the semantic and syntactic transformations of internalization are complete.

**Methodological Issues**

Because of its invisibility to external observation, inner speech has proved resistant to empirical research. Indirect methods include analysis of children's utterances during cognitive tasks, which has supported Vygotsky's ideas about the internalization, development, and self-regulatory functions of private speech. Direct methods for studying inner speech include self-report questionnaires (criticized for depending on potentially unreliable introspective processes) and thought protocols (involving the recording of what participants say when “thinking aloud,” a method limited by dependence on verbal formulation of subjective experience). Some limitations are avoided by experience sampling methods such as descriptive experience sampling, in which trained participants are interviewed about their experience immediately before the sounding of an electronic bleep.

Neurophysiological efforts to study the phenomenon began with electromyographical studies linking small activations of articularatory muscles to the subjective experience of inner speech. Neuroimaging studies have linked inner speech to activity in the left inferior frontal gyrus (including Broca’s area) and other brain regions. Such studies have been criticized for using ecologically invalid methods for eliciting inner speech and neglecting the possibility that inner speech will persist during baseline measurements.

**Theoretical Implications**

Inner speech figures significantly in a variety of psychological, neuroscientific, and philosophical
discourses. Its development in early childhood has been linked to a domain-general shift toward verbal mediation of cognition and behavior. Vygotsky’s ideas on the functional significance of inner speech increasingly figure in modern research into the executive functions, the heterogeneous set of cognitive capacities responsible for the planning, inhibition, and control of behavior. The dialogic nature of inner speech has been proposed to be important for capacities such as social understanding and creativity. Inner speech is also richly implicated in consciousness and self-understanding. Awareness of inner speech may provide important information for metacognitive understanding of one’s own mental processes, and the internal generation of verbally expressed higher order thoughts may play a part in conscious awareness. To the extent that it provides a running commentary on one’s actions, inner speech is potentially important in self-understanding and the articulation of views about the self.

Interest in inner speech has further been stimulated by its implication in disorders of consciousness. Auditory verbal hallucinations (associated with schizophrenia and other psychiatric disorders) are thought to stem from a misidentification of one’s own self-generated inner speech as having an external source, supported by findings of impaired source-monitoring in voice-hearers and commonalities in neural activation between auditory verbal hallucinations and inner speech.

Future Directions

A number of open research questions will potentially shape future research into inner speech. Conceptual challenges include providing greater specificity about important concepts such as internalization and charting areas of overlap and difference with other key processes such as subvocal rehearsal in working memory. Methodological challenges include improving the experimental elicitation of inner speech and further refining experience sampling techniques. Research on private speech in children as an empirically observable precursor of inner speech will benefit from improved methodologies for coding and classifying private utterances and empirical advances in the developmental cognitive neuroscience of the executive functions. The motivational and emotional aspects of inner speech, largely neglected to date, remain ripe for empirical exploration. Finally, the universality of inner speech remains an assumption that must be tested empirically and conceptually, which will in turn partly depend on further critical examination of the assumption of homogeneity. If there are several distinct varieties of inner speech, then many of our conclusions about its occurrence and psychological significance may need to be revisited.

Charles Fernyhough

See also Conversation and Dialogue; Language Development; Self-Consciousness

Further Readings


INTELLIGENCE, NEURAL BASIS

Intelligence, however defined, is in the brain. This entry explains why intelligence has a neural basis and how researchers are progressing toward the potential ability to enhance intelligence by manipulating biology. For research purposes, intelligence is reliably defined by tests of mental abilities that have a common general factor, usually designated g. Genetics play a predominate role in explaining individual differences in both general intelligence and brain structure. Because genes work through biology, there must be an important biological basis to intelligence, and this is the starting point for neuroscience investigations. Once salient brain properties are identified, such as regional gray matter volume or white matter integrity, further research can identify other biological and nonbiological factors that can influence these properties at any time during the life span. An understanding of the neurobiological factors related to intelligence will likely have implications for optimizing brain development, learning, and cognitive performance for children and adults.
Treatments for the low intellectual ability that defines mental retardation might be possible in some cases. Concern about Alzheimer’s disease has stimulated researchers to focus on the potential for drugs to increase learning and memory, two central aspects of intelligence. This raises a question as to whether any such successful drug, aimed at neurotransmitter systems and synaptic activity, could be used in persons without brain disease to increase general intelligence or even specific cognitive abilities related to but not the same as g (e.g., mathematical reasoning or musical ability). With these powerful motivations, neuroscience studies of intelligence are driven by increasingly sophisticated technology.

**Neuroimaging Studies**

Both clinical lesion studies in humans and experimental lesion studies in animals indicate that intelligence is represented throughout the brain in specific networks rather than in a single specific part of the brain like the frontal lobes. Neuroimaging studies in humans over the past 20 years have begun to identify these networks in some detail. Several functional imaging studies using techniques like positron emission tomography, functional magnetic resonance imaging, and the electroencephalogram, show that high scores on tests of general intelligence are related to decreased brain activity in some circumstances. This suggests that the efficiency of information flow throughout specific networks may be a key variable. Structural magnetic resonance imaging studies indicate that higher general intelligence scores are related to increased gray and white matter in specific brain areas and to regional cortical thickness, as well as to whole brain size. A comprehensive review of neuroimaging studies proposed that a parietal-frontal network underlies individual differences in general intelligence, and recent evidence supports this view. These studies raise the possibility of replacing traditional psychometric testing with “neurometrics,” meaning that intelligence can be defined and assessed by the size or shape of brain structures, or even by measuring the speed of information flow throughout salient networks with advanced techniques like magnetoencephalography.

Research results in the emerging field of “neurointelligence” already provide strong validation for psychometric tests of intelligence by showing scores are related to brain properties. Many issues, however, remain unresolved. Several neuroimaging studies, for example, indicate that the networks related to intelligence, and the way they develop through childhood, differ for males and females, even when matched on intelligence. This suggests at least two different brain architectures can lead to the same cognitive performance. There also may be separate brain networks for specific mental abilities like musical talent, artistic rendering, or mathematical calculation that have limited overlap with networks for general intelligence, as suggested by rare cases of autistic savants. Another important open question is whether differences in intelligence among people may be related to differences in how individual neurons function. For example, some people may have more efficient mitochondria (the part of the cell that produces energy) per neuron than other people. Characteristics of synapses and neurotransmitter systems differ among individuals as well. It would be odd if such brain differences had no effect on cognitive abilities or intelligence.

**The New Nature Versus Nurture Challenge**

The role of genes in many of these brain parameters is especially important. For example, both IQ scores and the amount of gray and white matter appear to have genes in common. The idea that intelligence may be under strong genetic control usually is interpreted to mean that intelligence is relatively fixed, as it is apparently not much influenced by environmental factors found within families. However, there already is evidence that genetic manipulation of specific receptors can lead to enhanced learning and memory in mice. Moreover, the Human Genome Project has resulted in rethinking simplistic genetic determinism because each gene can be expressed in multiple ways. There are biological and environmental factors, called epigenetic factors, which influence gene expression through mechanisms not yet understood. Individual differences in intelligence may be based on such interactions that likely influence structural and functional brain characteristics. These influences may differ according to age and sex, and they may help account for rising IQ scores observed over the past several decades. It may be possible to manipulate the genetic influence on intelligence by manipulating epigenetic factors once they are identified and the neuroscience is better understood. As a consequence of this complex challenge, it may be
possible to increase intelligence and other cognitive abilities in ways not now imagined.

Richard J. Haier

See also Flynn Effect: Rising Intelligence Scores; Heritability; Intelligence and Working Memory

Further Readings


**INTELLIGENCE AND WORKING MEMORY**

For much of the past century, many psychologists have sought to find a sort of Rosetta Stone that will allow for mapping the construct of intelligence on the one hand to more fundamental underlying psychological processes on the other hand. Although the construct of intelligence was developed early in the history of modern psychology, working memory is a recent conceptualization from basic experimental psychology. Attempts to link individual differences in intelligence to individual differences in working memory capacity have been controversial, but in the end, working memory appears to be just one, albeit an influential, component of a broader structure of intellectual abilities.

**Intelligence**

Two distinct traditions for the construct of intelligence developed over the course of the past century. The first is the concept of general intelligence (or g), which was proposed by Charles Spearman in 1904. In a groundbreaking review and set of empirical studies, Spearman determined that all manner of assessments of intellectual competence (e.g., grades in various school courses, performance on basic perceptual processes tasks) are positively correlated—a concept that has become known as “positive manifold.” That is, if an individual performs well on one measure of intellectual competence, that individual is more likely to perform better than average on any other measure of intellectual competence. Some of these correlations are relatively small, but among course grades, the correlations are more substantial. The implication of this pattern of correlations is that they define an underlying psychological factor that represents a general ability. The factor is present to a much greater degree in some measures (such as mathematics or language grades) than it is in other measures (such as the ability to discriminate among sounds that have very small differences in pitch). The factor was identified by Spearman as general intelligence (g). From this basic finding, Spearman argued that if one wants to estimate an individual’s intelligence, it doesn’t particularly matter what specific measures one chooses to administer or assess, as long as the measures have substantial “saturations” or loadings on the general factor. Adding up the scores on a series of tests may provide a reasonable estimate of the individual’s standing on the general intelligence factor. Spearman also argued that an individual’s general intelligence is “fixed”—that is, it does not change as one develops through childhood and adulthood; it is essentially fixed at birth.

Later developments in the tradition of Spearman’s general intelligence theory involved a series of investigations where researchers attempted to find particular tests that were highly saturated with general intelligence, in the hope of finding just one test that could accurately provide an assessment of an individual’s intelligence. After several attempts, Spearman and his colleagues settled on a test of abstract inductive reasoning, called Raven’s Progressive Matrices test, in the 1930s. Since the 1930s, there have been many studies of individual differences in performance on the Raven test, and in the 1980s forward, it has been a central focus of experimental psychologists who seek basic information processing tests that have high correlations with performance on the Raven test. The underlying theorizing of these investigators is that if a task that
The concept of working memory was introduced by Alan Baddeley in the 1970s. The concept was rooted in prior efforts to classify and understand, from an experimental psychology perspective, the basic building blocks for cognitive functioning. Earlier theories and research in the 1950s and 1960s suggested that people have separate storage systems for short-term memory (which lasts for seconds if the information is not consciously attended to) and long-term memory (which is a more permanent storage system). These storage systems accounted for a wide range of experimental phenomena, but failed to account for how people manipulate information in conscious attention and for other aspects of the cognitive information processing system. Baddeley and his colleagues proposed instead that the information processing system is made up of a central executive and two subsidiary systems. The central executive is involved in planning and manipulating information in conscious attention, while the phonological loop (the first subsystem) operates on speech-based information, and the visuospatial sketchpad (the second subsystem) operates on visual images.

From this theoretical perspective, several tasks have been developed that allow assessment of different aspects of the central executive and subsystems. The accuracy of the intelligence assessment for Binet and his colleagues was determined by how well the scores on his scales correlated with academic success or failure.

The Binet approach for assessing intelligence ultimately became the standard for future applied assessments. Modern intelligence tests, such as those of David Wechsler and the Stanford-Binet, have many of the same underlying components as Binet’s original scales published in 1905. However, because the approach adopted by Binet and his followers specifies that intelligence is both broad and context-dependent (in the sense that culturally bound knowledge and skills are important components of cognitive functioning), this approach is not generally amenable to a search for a single or a few underlying processes that lead to individual differences in intellectual functioning. As a result, there are relatively few investigations that have attempted to link this kind of intelligence measure to basic experimental processes.
Intelligence and Working Memory

other hand. From a theoretical perspective, a correlation close to 1.0 between such measures would indicate that these are closely related constructs and, in turn, might suggest that the limiting process for high intelligence is the ability to store and manipulate items with working memory. From a practical perspective, if the correlations between these measures are close to 1.0, then, for most intents and purposes, one could administer a test of working memory and obtain an accurate assessment of intellectual ability. Some investigators boldly proposed that working memory and intelligence were isomorphic, that is, that individual differences in working memory capacity were the same as individual differences in intelligence. Both such claims and the basis for the claims were controversial, for two main reasons.

The first controversial aspect of these claims was that they were predicated most centrally on correlations between measures of working memory and intellectual ability measures of the kind that were developed to assess Spearman’s g; that is, measures of abstract reasoning such as the Raven test that were mainly devoid of cultural or verbal content. These tests capture only one aspect of general intelligence, that is, the part of intellectual functioning that is associated with nonverbal content. In addition, these tests were not highly correlated with academic success, as are the Binet-inspired tests, and as a result, the association between working memory capacity and aspects of intelligence that have clear real-world implications was not known.

The second controversial aspect of these claims was a function of how working memory capacity was measured. In many studies, researchers used a technique called “extreme-groups” study. That is, instead of testing a group of individuals with a wide range of talent (such as would be obtained if one randomly sampled from the population at large), these researchers would test a large number of individuals and eliminate all but those with extremely high and extremely low scores on the working memory test. Then, they would look to see if these extreme groups also had differences in mean performance levels on intellectual ability measures. The problem with this kind of research design is that it overestimates the differences between individuals, sort of like looking at basketball performance of extremely tall and extremely short groups of people.

To provide a better sense of the actual relationship between measures of working memory on the one hand, and measures of intellectual abilities on the other hand, some researchers conducted a meta-analysis. A meta-analysis entails collecting all of the studies that have been conducted on the topic, categorizing each different working memory test and each different intellectual ability test, and then averaging the different results to get a robust estimate of the overall relationships between the measures. This technique, developed in the 1980s, is a useful way of summarizing multiple studies of the same general topic, and it minimizes a number of biases and errors that can be found in individual studies, especially when the studies have small samples or few measures to assess the underlying constructs.

In the final meta-analysis, the results suggest that many measures of working memory are indeed correlated with measures of intelligence, but at a much more modest level than was claimed by those who hoped that working memory represented the essence of intellectual functioning. The estimated correlation between the two constructs was about .50, meaning that working memory accounts for about as much variance in broad measures of intelligence as other, more traditional, ability measures, such as vocabulary, numerical reasoning, spatial visualization, and so on. Individual differences in working memory are undoubtedly related to individual differences in intelligence, but it appears clear that working memory is not the single limiting factor for intellectual functioning.

Phillip L. Ackerman

See also Flynn Effect: Rising Intelligence Scores; Intelligence, Neural Basis; Working Memory

Further Readings


INTENSION AND EXTENSION

Intension and extension are terms of philosophical art in the philosophy of language, linguistics, and the philosophy of mind. Unfortunately, these widespread terms of art are used somewhat differently by different theorists. Perhaps the best way to proceed is to describe orthodox uses of the terms intension and extension and then to note why some writers use them differently.

Extensions

Here is a basic package of views about what extensions are. The extension of a predicate is the set of objects that it applies to. The extension of “is a rabbit” is the set of rabbits, for example, and the extension of “runs” is the set of running things. The extension of a name is the thing it is the name of. The extension of “George W. Bush” is a certain ex-president, for example. The extension of a sentence is a truth-value: normally there are thought to be two truth-values, True and False. Sometimes it is thought that an extension is the thing, or things, “out there” in the world that are picked out by pieces of language.

In the years since Gottlob Frege first presented this package of views about extensions, some theorists have offered slightly different accounts of the extensions of these expressions. Some people are suspicious of sets, or of giving sets a role in the meanings of ordinary pieces of language. They might interpret the extension of a predicate as the things, collectively, that the word applies to (the rabbits are the extension of “rabbit”). Other people think that the extension of a predicate is only the set of things that the predicate applies to in this possible world, and reserve the set of all actual and possible things the predicate applies to for the role of an intension. (See the suggestion by David Lewis in 1970 about extensions.)

As well as disputes about what sort of things should be considered to be the extensions of predicates, names, and sentences, there have also been theories that extend the notion of extension to other parts of language. Quantifier expressions (“all cats,” “most sailors,” etc.) are assigned extensions in some kinds of generalized quantifier theory. Descriptions (“the first King of England,” “a cat in my garden”) are sometimes assigned extensions, either by treating them as a kind of quantifier expression or by treating their extensions as the things they apply to. Extensions could be attributed to other pieces of language as well.

Intensions

Intensions, on the other hand, are often treated as features of meaning that determine the extensions of words. The intension of a predicate might be a concept or something that describes the things that fall under the predicate. Or it might be a property, such that things fall in the extension of the predicate if they possess that property. Or it might be a function from possible worlds to extensions—given the relevant possible world as an input, it delivers the extension of the expression as an output. On this conception of the intension of predicates, these intensions would in effect code information about which things would have which features, given different alternatives to this world.

A similar range of options exist for names. The intension of a name might be some concept or description that picks out the thing it is the name of. Or it might be a property such that only one thing (the thing named) has that property. Or it might be a set of properties that together uniquely characterize the object named. (This approach comes from the work of Richard Montague.) Or it might be a function from possible worlds to objects. Finally, the intension of a sentence might be a full thought, or a description of the circumstances under which it is true. Or it might be a proposition that, together with how things are, determines a truth-value. Or it might be a function from possible worlds to truth-values—given a world and a sentence, the function yields the extension (i.e., the truth-value) of the sentence.
There is significant variation in what theorists take intensions to be. The most prominent theory of intensions, however, treats intensions as being ultimately defined using possible worlds and set theory. Possible worlds are alternative ways things could have gone. One reason they are invoked is to show not just what words do apply to but how they would have applied if things had gone differently. Montague was perhaps the chief pioneer of the tradition of taking intensions to be built up from possible worlds and the contents of those worlds, and this treatment of intensions is now common among philosophers of language but especially among semanticists in linguistics. One way to do this is to treat intensions as functions from possible worlds to extensions: in effect, to treat the intension as a rule for assigning pieces of language various extensions given different specifications of how the world turns out. So the intension for “is a rabbit” takes a world \( w \) as input and delivers a set of possible objects as an output—the set of objects that are rabbits in \( w \). The intension of a sentence is then a function from worlds to truth-values. We could equivalently represent such an intension as the set of worlds where the sentence is true, since the set of worlds where the sentence is true is the one that is mapped to the truth-value True by a function from worlds to truth-values. There are a number of proposals for assigning intensions to names. One of the simplest would be a function from worlds to possible objects in that world (i.e., the object named in that world). An even simpler intension of a name would be just the object itself—we could maintain that the name names that very object, no matter what possible world we consider.

Contemporary semantic theories assign intensions to many other categories than proper names, predicates, and sentences. For example, in the tradition of defining intensions as set-theoretic constructions from possible worlds and individuals, one way to specify the intension of the word and is as a function from two sets of worlds to one set of worlds, in particular, the function that takes two sets of worlds and delivers their intersection. To see how this works, notice that we can treat the proposition expressed by a sentence as the set of worlds that are as the sentence says things are. When we have two sentences and connect them with and, the resulting sentence is true when, and only when, both of the initial two sentences are true. In other words, the conjunction is true in the intersection of the two sets of worlds where each of the conjuncts is true.

One aim of many theories of meaning is to provide a compositional semantics for linguistic expressions. Specify the meanings of the expressions, and specify the way they are put together, and the theory should predict the meaning of the larger expression produced. The goal of semantic theories like Montague’s is to assign intensions to every part of language and then formulate general principles that allow us to determine the intensions of complex expressions from the intensions of their constituents. Some theories of meaning try to get by without postulating intensions, and instead provide rules for determining the extensions of sentences by specifying the extensions of some of the sentence’s constituents. Such theories are much less likely to assign extensions to every piece of language used in the sentence, however, because there seem to be insurmountable problems for determining extensions for expressions such as necessarily.

Extensional and Intensional Contexts

A traditional place extensions and intensions have played a role in philosophy of language is in specifying the difference between an extensional context and an intensional context. An extensional context is a position in a sentence where coextensive expressions can be substituted without changing the truth-value of the overall sentence. For example, “Jon” occurs in an extensional context in “Jon hugs Mary.” Any other expression with the same extension as “Jon” could be substituted there. For example, if one of Jon’s other names was “Jonathan Dodgson,” then the sentence “Jonathan Dodgson hugs Mary” would have to have the same truth-value as our original sentence “Jon hugs Mary.” If definite descriptions such as “the nicest postman in Grimsby” have extensions (a controversial matter), and that expression happened to be coextensive with “Jon,” then we could substitute it and be sure of getting a statement with the same truth-value.

Some places in sentences, however, are intensional. Not just the extension of the expression in that place matters to the truth-value of the claim, but the intension does as well. Take our previous example again, and suppose “Jon” and “the nicest postman in Grimsby” have the same extension (they both pick out Jonathan Dodgson). “Mary believes
Jon is a postman” may come apart in truth-value from “Mary believes the nicest postman in Grimsby is a postman.” Mary may not have talked to Jon about his work, but she is still capable of working out that the nicest postman in Grimsby must at least be a postman. This shows that “Mary believes ___ is a postman” can be an intensional context. Predicates can also appear in intensional contexts. Suppose, as seems plausible, that all and only giant pandas are vegetarian bears (i.e., “is a giant panda” and “is a vegetarian bear” have the same extension). A zookeeper might say about the zoo’s brown bear, “If we train Growley to eat only tofu, Growley would be a vegetarian bear.” It would not thereby be true that “If we train Growley to eat only tofu, Growley would be a giant panda.” “If . . . then . . .” constructions, therefore, can be intensional contexts.

### Intension, Hyperintension, and Intentionality

Places in sentences are called *hyperintensional* when substituting expressions with the same possible-worlds intensions can produce a difference in truth-value. Belief contexts are like this: In the previous example, “Jon” and “Jonathan Dodgson” have the same *intension* according to many theories, but Mary might not believe that Jon is Jonathan Dodgson, if she knows about Jonathan Dodgson as the author of her favorite novels but Jon as the local postman. On the other hand, she does believe that Jonathan Dodgson is Jonathan Dodgson, so we cannot substitute the two names in our belief reports without affecting truth-value. Hyperintensional contexts may signal that we need something more complex than possible-worlds intensions for a theory of meaning.

It is important to distinguish intensionality (with an *s*) from intentionality (with a *t*), especially since they are often used about similar subject matters. Intentionality (with a *t*) is perhaps best glossed as “aboutness,” or as the ability of mental states in particular to represent other things. Belief states are intentional, for example, because they have contents: a belief that John is hungry is about John and is also perhaps about being hungry. Franz Brentano famously claimed that all mental states were intentional, although this is not obviously true. Perhaps being tired, for example, is at least in part a mental state, but is not obviously about anything. We are not normally tired *at or about* anything at the end of a long day, even when it is more mental tiredness than physical tiredness.

Talk about intentionality is often intensional (with an *s*). Belief reports are often intensional, as are many other reports of mental states. Other mental states are “directed” but seem to require description in ways that create non-extensional contexts. For example, the sheriff might hunt for Robin Hood but not hunt for Robin of Locksley (he may even have enlisted Robin of Locksley’s assistance to help him hunt for the dangerous outlaw). Or at least this seems true for one use of “hunts for.” The expression “hunt for” seems to create a non-extensional context and picks out an intensional state of the sheriff, or at least a partially intentional state of the sheriff, as hunting might involve an activity as well as a mental state.

The home of the concepts of extension and intension is the theory of language. However, they are also important in the philosophy of mind: Mental representations may well also have extensions and intensions, and theorizing about these seems indispensable to theories of mental content. How the notions of extension and intension apply to mental representation depends, of course, on what mental representations are like. If mental representation is via a language of thought, the distinction drawn for languages can just be applied as is. If mental representation works in other ways, the account may have to be different. Still, theories of mental content will probably need to take account of intensional phenomena as well as extensional ones: The representational character of a complete thought is not exhausted by its truth-value, nor can the mental representations of objects be exhaustively characterized by specifying which object is represented. It matters *how* objects are represented, and those differences in representation are analogous to the way that pieces of language that refer to objects can have the same extension but differ in intension.

*Daniel Nolan*

### See also

Concepts and Language; Descriptions; 
Representational Theory of Mind

### Further Readings

Frege, G. (1952). *On sense and reference*. In M. Black & P. Geach (Eds.), *Translations from the philosophical...*
**Intentionality of Bodily Sensation**

A mental state or event is said to have intentionality when it is about, is directed toward, or represents, something. Some philosophers claim that bodily sensations, such as itches, tickles, tingles, and pains, lack intentionality. Among those who defend the claim that bodily sensations do possess intentionality, there is disagreement over what they are directed toward, and there is disagreement over the account that should be given of the manner in which they are directed toward their objects. This entry will briefly explore these issues.

**Puzzles About Bodily Sensations**

Reports of certain bodily sensations seem to involve the use of transitive verbs whose object is some sensible quality of an external object or event, such as temperature or pressure—for example, “I can feel the warmth of the cup.” Other reports appear to involve the use of intransitive verbs—for example, “My foot is itching.” Disagreements over whether bodily sensations have intentionality are usually concerned with the latter, “intransitive” bodily sensations. Some have argued that although sensations like itches, tickles, and pains, may be caused by events in one’s body, such mental occurrences are not directed toward anything. It is suggested that although we can distinguish between a perceptual (e.g., visual) experience and what that experience is an experience of, this sort of distinction is inapplicable in the case of bodily sensation.

A common objection to this position is that it fails to accommodate the felt location of bodily sensations. For example, some argue that the phenomenological difference between an itch on your nose and an itch on your back is best accommodated by holding that such bodily sensations involve experiences that are intentionally directed toward different locations. If a bodily sensation, such as a tickle, is identified with something that is felt to have a location on some part of one’s body, then arguably the bodily sensation is not itself an intentional mental state or event. Rather, the experience of it is. To mark this distinction, two distinct sensation terms are sometimes introduced: one that refers to the sensation as intentional experience (henceforth, SE) and one that refers to the object it is directed on (henceforth, SO).

**Accounts of the Intentionality of Bodily Sensation**

For those aiming to provide an account of the intentionality of the experience of located bodily sensation, three central concerns are the following: (a) to provide an account that accommodates the distinctive manner in which a bodily sensation feels to be located; (b) to provide an account of what the subject is aware of at the relevant location, for example, whether it is simply some part of her body, or something obtaining or occurring at that part of the body; and (c) to provide an account that can accommodate the affective dimension of certain bodily sensations, for example, the negative affective component of pain.

**The Feeling of Bodily Sensation as Located**

It is often suggested that when one feels a bodily sensation as located, the relevant location is primarily given as the location of some part of one’s body. For example, in the case of “phantom limb” sensations, experienced by some amputees, it does not feel to the subject as though there is a pain at some place in midair. Rather, it feels to the subject as though her body is located at that place, even though the relevant body part no longer exists. In light of this, some claim that one’s awareness of a sensation as located is itself a form of bodily awareness—a form of awareness of one’s body, or some part of it. Under this approach, providing an account of the intentionality of located bodily sensation is connected with the task of explaining the sense in which one is aware of one’s body “from the inside.”
What the Subject Is Aware of at the Location

According to some accounts, when one feels a sensation as located in some part of one’s body (e.g., a stabbing pain in one’s knee), one has an experience that is not only directed toward some body part but is also directed on a subjective, mind-dependent entity. For example, the suggestion is that although your knee seems to you to be the kind of thing that can exist independently of your awareness of it, the stabbing pain (SO) that you feel to be there does not. Connected with this is the suggestion that, although other subjects may come to know that you are feeling a stabbing sensation in your knee, there is something odd in the idea that they might come to know this through feeling it. So although your knee is a publicly observable object, the pain you feel to be there is not. Some claim the relevant entities are mind-dependent objects (sense data); others claim that they are subjective properties of body parts; and yet others claim that, when we feel such sensations, we project mind-dependent qualities onto parts of the body.

According to an alternative, “representationalist” approach, although the intentional experience (SE) of a bodily sensation is a subjective mental occurrence, what it represents as obtaining at the body part is not. One’s experience of a bodily sensation has an intentional content that represents some mind-independent, physical condition of the body. Some who take this approach suggest that our ordinary talk about bodily sensations tends to focus on the subjective intentional experience, rather than the mind-independent condition it represents, because the affective quality associated with the experience (e.g., the negative affective quality associated with pain) is so salient.

Affective Dimensions

When a subject feels severe pain in some part of her body, she might be described as “being in pain.” Some suggest that this psychological condition has emotional and motivational components that cannot be accommodated by simply appealing to the fact that the subject is having a conscious experience that represents some bodily condition. Although there may be a sensory-discriminatory aspect to pain experience—for example, concerning the location, intensity, and quality of the pain—we also need to appeal to distinct emotional-cognitive components of pain experience in order to explain how the experience can come to signify for the subject the threat of damage to her body. Others argue that the negative affective component of a sensation like pain is an aspect of the representational content of one’s experience. For example, in feeling severe pain in some part of one’s body, a bodily condition, such as tissue damage, is represented by the subject’s experience as bad or as unpleasant, and the suggestion is that subjects have the normal cognitive and emotional reactions to pain that they do because their experience represents the bodily condition in this way.

Matthew Soteriou

See also Causal Theories of Intentionality; Intentionality of Emotion; Mind-Body Problem; Representational Theory of Mind

Further Readings


Intentionality of Emotion

Intentionality refers to the ability of mental states to be about, be directed at, or represent objects, properties, and states of affairs in the world. Just as it is disputed whether all mental states are intentional, so it is not settled whether all emotions have intentionality, but it is generally accepted that most emotions do and that this is one central way of distinguishing them from similar but non-intentional states such as moods. I am angry at someone, in love with my wife, afraid of the shark, worried by the shouting, happy about winning the lottery, surprised because of the kind act. Moods such as depression, in contrast, do not seem to be directed at anything at all, or at least at no particular or a less specific object. Providing an account of the intentionality of emotion is seen as one of the key elements in understanding the nature of emotion, and many of the differences between competing philosophical theories of emotion can thus be seen to hinge on different ways of explicating emotional intentionality.
The nature of intentionality is, however, complex, obscure, and subject to a great deal of philosophical debate, so it is unsurprising that, given the many and varied dimensions of emotional phenomena and the raging debates about defining emotion, the subject of the intentionality of emotion cannot avoid inheriting and multiplying this difficulty and complexity. As such, the reader should be aware that many of the issues discussed in this entry are hotly contested.

This entry is structured as follows: The first part provides an overview of some of the main issues concerning the intentionality of emotion, including the nature and range of the intentional objects of emotions. The second part examines the view of emotional intentionality inherent in cognitive and feeling theories of emotion and notes some of the problems facing such views. The third and final part examines some perceptual theories of emotion, the various views of intentionality arising from such theories, and some of the problems and challenges confronting these, such as explaining the role of the body and feelings in such intentionality.

**Intentional Objects of Emotion**

One central obstacle to capturing the intentionality of emotion in any unified way is that the intentional objects of emotions form an immense and varied range, including actions, events, values, states of affairs, particular objects, properties, propositions, and bodily states. If I am afraid of the lion, my fear is directed at a particular object, the lion, and perhaps also at the situation confronting me. We might also capture this by saying that I am afraid that *the lion will eat me*, thereby giving the emotion a propositional object (in italics). Whereas some emotions, such as anger, may be directed at either propositional objects or particulars, others, such as love, may require specific objects as their focus, and yet others, such as regret, may require propositional objects. Moreover, certain social emotions such as pride, jealousy, and shame, often involve a number of different intentional objects, partly in virtue of their self-reflexivity. If I am ashamed of performing a certain action, this may involve negatively evaluating both the action performed—for example, as shameful, or more specifically as morally or socially bad or unacceptable—as well as myself, as perhaps unworthy or in some respect dis-valuable.

Yet further complexities become apparent when it is noted that the intentional objects of emotions may not just be perceived as present, but may be absent or even merely imagined. Whereas some emotions, such as anger at *p* or love for *p*, might seem to require belief in the existence of *p* (or where *p* is a proposition, the truth of *p*), others, such as fear or hope, do not. Hope, for example, may take a propositional but nonexistent object, as in “I hope that the green dragon will not eat the hobbit.” Imagining or construing some object as disgusting may move one to feel disgust as if the object were genuinely perceived to be disgusting, even when it is not believed to possess disgusting properties.

Both philosophers and psychologists maintain that there is thus a close connection between emotions and evaluative judgments or appraisals. Debates about the right conception of emotion center here around whether the relevant appraisals cause the emotion or are constitutive parts of the emotion. So, one might say that the perception of the lion, and the realization that it poses a threat to me, causes my fear, where the actual emotional state is conceived as something independent of this overall evaluative appraisal; or one might say that this appraisal just is, in part, the emotional state of fear itself. What answer one gives to this will depend, in part, on whether one thinks of emotions as cognitive or noncognitive states.

Emotions are not just directed at particular objects, actions, and states of affairs, but at more specific features and properties of these that explain the specific emotional responses. My fear of the lion, for instance, is explained by my perceiving his big teeth and menacing gait as frightening and hence my construing the situation as dangerous. It is the danger or dangerousness that explains why I feel fear rather than surprise or boredom. In this way, philosophers sometimes distinguish between material (or particular) intentional objects—the lion and its big teeth—and formal (or proper) intentional objects—for example, the danger. The latter depend on the former and generally comprise evaluative properties, or they essentially involve evaluations. It is not the lion as such that is the object of my emotional state, but the lion *construed* as dangerous.

In fact, the purported distinction between material and formal objects is open to some debate, but it is supposed to capture the central role that the latter play in defining and individuating the emotions directed at
them. So fear, for example, may be defined partly as a response to the dangerous or threatening, and this is in part what distinguishes it from other emotions, such as anger. Some philosophers and psychologists have thus attempted to develop taxonomies of emotions based on their formal objects and the “paradigm scenarios” or “core relational themes” that contain them. Sadness is a reaction to loss, anger a reaction to the offensive, fear a response to danger, and so on.

Formal objects are also generally held to play a central role in explaining and justifying emotional responses. My fear of the lion is explained or made intelligible by the danger the animal poses, and only if an object possesses a certain value, such as dangerousness, is fear an appropriate, warranted, or correct response. Formal objects thus provide not merely criteria of intelligibility, but criteria of correctness and rationality for emotions insofar as they give reasons for the subject to feel specific emotions rather than others. It is worth noting that if the formal objects of emotions are taken to be value properties, then one’s metaphysical view of the status of such properties may have ramifications for one’s views about the rationality of emotions and the nature of their intentionality.

Over and above these detailed issues and disputes, it is widely accepted that the intentionality of emotion comprises both an inward focus, on our own bodily changes and phenomenological state, and an outward focus, onto the range of objects in the world at which our emotions are directed. On the one hand, therefore, the intentionality of emotions can be understood in terms of subjects representing the objects of their emotions, and this helps to distinguish the objects and causes of emotional responses, as these can come apart. For example, I may react with anger to some trivial, nonoffensive remark you make, where the real cause of my anger is a deep-seated mood of depression from which I am suffering, and yet this is clearly not represented by me as the object of my emotion. On the other hand, however, there appear to be certain nonrepresentational intentional properties of emotions, such as somatic changes and bodily “feelings,” often explicated in terms of certain biological, evolutionary functions. A challenge for emotion theories is to reconcile or combine these two different aspects of emotional intentionality.

Cognitive Theories and Intentionality

Many of the features of the intentionality of emotion listed earlier explain, in part, the shift from what are frequently called “feeling theories,” prominent for much of the first half of the 20th century, to the “cognitive theories” that replaced them and which have, in turn, held sway until relatively recently. Supported by work in empirical psychology, feeling theories hold that what is essential to the nature of emotion are certain bodily changes, and they either identify the emotion with such changes, with the feeling or perception of such changes, or with a wider process involving both the changes and perception of such changes. The intentionality of emotion on these views is at best restricted to the bodily changes of which we are, through some sort of feeling, aware.

Such theories appear to provide an inadequate picture of emotional intentionality, failing to individuate the vast range of emotions via a relatively limited range of bodily feeling and changes, and essentially neglecting the intentional relationship of emotions to the world. I am afraid of the charging lion, not of my bodily state. Concomitantly, these theories face difficulties in explaining the rationality of emotions.

Cognitive theories, in contrast, have sought to model the intentionality of emotions on paradigmatic intentional states, such as beliefs and desires. A typical cognitive model holds that an emotion such as fear can be understood as involving the belief that something is dangerous and the desire to avoid this danger; anger consists in the belief that I have been wronged and a desire to retaliate, and so on. Or, on a more reductive account, fear just is the judgment that something is dangerous, and anger just is the judgment that I have been offended. In either case, the belief that an object is dangerous is thus a necessary condition for being afraid and, perhaps coupled with desire (to avoid the danger), forms a jointly sufficient condition for fear. Thus, emotions involve beliefs about their material and formal objects, and the rationality of emotional responses can be understood in terms of the truth and falsity of such beliefs.

Many philosophers, however, have objected to this model both as a way of accurately capturing the nature of emotions and as a way of explaining their intentionality. Although they do not deny that emotions frequently involve certain beliefs and desires, these philosophers argue that emotions do not seem reducible to combinations of these states or to be fully captured by reference to them. Beliefs do not seem to be necessary for emotions, for these may exhibit cognitive impenetrability—I remain afraid of the dog even though I do not believe the dog is dangerous. Neither do they seem to be sufficient—I may believe the lion is dangerous without feeling fear,
I may believe the death of a loved one is a great loss without feeling sadness. Doubt has also been cast on the close link between emotions and desires apparently posited by cognitive theories. Feeling pride or admiration, for example, do not seem necessarily to involve any particular desire.

In light of such considerations, there has been a steady but noticeable convergence of views to the effect that the belief-desire model of emotions is far too impoverished a way of understanding our emotional engagement with the world. In particular, it seems obvious that emotions typically have a phenomenological flavor that beliefs (and perhaps even desires) lack, and that some emotions have distinctive bodily changes associated with them. That is, it feels like something to be in an emotional state, as opposed to a belief-state that doesn’t feel like anything and hence does not have any distinctive phenomenology. Arguably, this phenomenology of feeling and the associated bodily changes are not just contingent parts of emotions accompanying the underlying beliefs and desires; rather, they constitute essential parts of the emotional response. It has thus been maintained by some that the intentionality of emotions cannot be reduced to the intentionality of other states such as beliefs and desires, but it remains unclear what role, if any, “feeling” should be thought to play in characterizing emotional intentionality.

**Perception, Feeling, and Intentionality**

Partly in an attempt to explain some of these phenomena, many recent theories have stressed important analogies between emotions and perceptions and a number of “perceptual theories” have been put forward. The label is a relatively loose one, encompassing as it does a host of quite different looking models of emotions and their intentionality. The appeal of such theories is that they seek to capture the way in which emotions, like perceptions, have a certain phenomenology: that they often have an image-like quality; that their triggering is, to a large extent, automatic or involuntary; and that, on the whole, they are belief independent.

Peter Goldie, for instance, argues that the bodily feelings present in many emotions—those that have a distinctive physiology—have intentionality in the sense that these feelings are directed toward one’s body as being a certain way, and as such they provide prima facie reasons for believing that one is experiencing a particular emotion. In addition, he argues that emotions have a distinctive intentionality involving what he calls “feeling toward”; roughly, a kind of thinking of *x with feeling* that is directed intentionally at objects in the world as being a certain way or having particular features. “Feeling toward,” unlike belief, may be subject to the will and has a distinctive phenomenology. Feeling afraid of the lion involves having feelings of fear toward the lion, where this sort of intentional state cannot be analyzed in terms of unemotional attitudes.

Although it recognizes the distinctive, complex, and rich nature of emotional intentionality, it might be argued that this notion of “feeling toward” nonetheless remains a little too obscure to helpfully explain such intentionality. But further perceptual theories of emotion promise to flesh out the relationship between bodily feelings and intentionality in a way that does justice to the phenomena.

Jesse Prinz, drawing heavily on contemporary psychological theories of emotion, has developed perhaps the most comprehensive attempt to reconcile the inward and outward aspects of emotional intentionality. Emotions, Prinz holds, are “embodied appraisals.” They are individuated by those situations that reliably elicit them in virtue of possessing what he calls “core relational themes”; namely, relational properties holding between us and our environment that bear on our well-being, for example, dangers, losses, offenses, achievements. A dangerous situation, for example, will reliably elicit fear. Further, Prinz argues, mental states get their intentional content in virtue of having the function of being reliably caused by something. If emotions are perceptions of bodily states, therefore, and those bodily states are reliably caused by the instantiation of core relational themes, then our perceptions of the body may also represent those themes. When confronted with a dangerous situation, the appraisal of danger elicits those bodily changes that partly constitute fear. The perception of such changes then represents the core relational theme of “danger” because it is reliably elicited by dangerousness. Thus, Prinz claims, perceptions of bodily feelings have formal objects—for example, danger—and conditions of correctness, and the intentionality of emotions directed at inward bodily states is correlated with their intentional relation to the world outside.

There are some, however, who doubt that such a close link can be forged between emotion and perception. Whereas the content of our perceptions appears to be straightforwardly causally connected to what is the case in our environment, emotional
intentionality essentially involves and depends on an extremely rich set of background values, needs, aspirations, beliefs, desires, expectations, and self-conceptions. Emotions do not so straightforwardly track the way our environment is.

Rather different perceptual models, such as those put forward by Ronald de Sousa and Robert Roberts, seek to explain these differences—and thereby provide a more accurate picture of the intentional “feelings” present in emotional intentionality—by holding that emotions are not best thought of as a straightforward type of perception but rather as “construals” or ways of seeing or attending to the world. The function of emotions on such views is partly to make certain features of situations, actions, and objects salient, to foreground them in our appraisals and to direct our attention to them in preparation for judgment and action. The intentionality of emotion then involves a complex interaction between imagination, belief, perception, and feeling, as well as various motivating states and values of the agent. For example, to construe a situation as dangerous, I need to perceive it as bearing a particular, threatening relation to my various values, desires, and interests, and my feeling of fear will in turn help me to focus on those features of the situation that are most relevant to its danger.

Cain Samuel Todd

See also Emotion, Psychophysiology of; Emotion, Structural Approaches; Emotions and Consciousness; Intentionality of Bodily Sensation; Rationality of Emotion

Further Readings


INTERGROUP CONFLICT

Intergroup conflict emerges from three elements of human psychology: group perception, group identification, and group threat. All three are required for substantial conflict and especially for conflict that includes violence: interstate war, genocide, ethnic cleansing, pogroms, riots, and terrorism.

Group Perception

Intergroup conflict presupposes perception of at least two groups. This is easy when two teams meet on a sports field, with each team member marked by the team colors. Often, however, group perception is not so obvious. How do Irish Republicans know who is a Loyalist in Northern Ireland? How did Hutu know who was Tutsi in the 1994 genocide in Rwanda? How did “real” Cambodians know who were the “Cambodians with Vietnamese minds” to be exterminated in the killing fields of 1975–1979?

Interrmarriage, conversion, and migration produce many individuals who cannot easily be coded in an either-or fashion. Is the son born to Muslim Algerian parents in France seen as French or Algerian? This is not a matter of what passport he holds. How do most French people see him? How do Algerians see him? How do other French citizens of Algerian origin see him? Not only are group boundaries fuzzy, but they can change over time. In the 1800s, Irish immigrants to the United States did not count as “White”; today their descendents are undeniably White.

The puzzle is why we see these fuzzy boundaries as sharp and well defined. When someone talks
about Loyalists, or Tutsi, or Jews, only a few social scientists are likely to say that, well, actually, we don’t know who you are talking about. Groups made up of uncounted millions, separated by porous and changeable boundaries, sometimes with little difference in physical appearance (Irish Republicans and Loyalists, Cambodians with and without Vietnamese minds) appear with no less clarity than two teams on a football field. The more intergroup conflict moves toward violence, the sharper and more natural group boundaries appear to be.

The capacity to see a group composed of millions of diverse individuals as a single entity, a single actor, is an aspect of human psychology that rivals the capacity for language. A child does not have to be rewarded for learning language and does not need explicit instruction from adults to learn language. Similarly, humans seem prepared to learn about groups without external reward or explicit instruction, and this group-making tendency extends to seeing ethnic, national, and religious groups with no less clarity than football teams.

**Group Identification**

The next puzzle has to do with caring about groups, part of a broad human capacity to care about others. Caring about family and friends seems so natural that it does not draw much attention, although the borders of who counts as “family” or “friend” can be as fuzzy as the border of nation and ethnicity. But we care also about much larger groups: our ethnic group, our religious group, our country. Unlike family and friends, these groups are too large to be known as individual persons; these are groups of strangers. Nevertheless, we often care for these anonymous groups to such a degree that we are ready to give up friends, family, and even our own lives for the welfare of the group. Sometimes we even contribute to the support of groups we are not part of, such as a football team or Darfur refugees.

In short, humans often care about groups in a way that cannot easily be explained as self-interest. We do not lose track of self-interest when we sacrifice for our team, our ethnic group, or our country. Rather, we choose to put group interest above self-interest for the groups we care about. This capacity to care about groups, when joined with our far-reaching preparedness to see the social world divided into groups, opens the door to intergroup conflict. All that is required is perceived threat.

**Group Threat**

Group conflict emerges over three kinds of perceived threat: threat to material resources, threat to status or honor, and threat to security. Threat to security includes threat to the material existence of a group (e.g., the kind of mass murder that Rome brought to defeated Carthage) as well as threat to the purity or essence of a group (e.g., the kind of contamination Hitler feared Jews would bring into the German volk or that Pol Pot feared Cambodians “with Vietnamese minds” would bring into “authentic” Cambodians). These threats usually overlap to some degree. Status and honor are often associated with access to material resources and security from attack. Conflict over resources is often seen to engage issues of honor and long-term security. Purity is often seen as the key to strength and honor.

Studies of social movements suggest that to turn shared threat into collective action requires mobilization of resources (people, skills, money), political opportunity (costs and benefits of action, including potential allies), and a cognitive “frame” that identifies the bad people responsible for threatening the group and the good people who should act to respond to this threat. These requirements apply no less to governments than to groups and movements that seek to challenge governments, although governments usually have superior power for mobilizing resources. Collective action against a common threat has powerful effects on in-group dynamics: idealization of in-group values, increased respect for in-group leaders, and increased punishment for deviation from in-group norms. These effects are so reliable that group leaders have been known to foment intergroup conflict in order to profit by the rallying effect at home. Especially for violent intergroup conflict, the dynamics of conflict often play out quite differently for leaders, perpetrators of violence, and the mass of group sympathizers and supporters.

**Conclusion**

Most intergroup conflict is not violent. Countries, corporations, political parties, and religious groups usually compete without violence. Around the world, hundreds of minority ethnic groups offer only peaceful challenges to a dominant majority group; at any given time, only a few ethnic groups are involved in riot, rebellion, or terrorism. Escalation of conflict to violence usually occurs over a considerable period
of time, in a sequence of action and reaction that sharpens boundaries, raises group identities, and increases threat perception on both sides.

Clark McCauley

See also Collective Action; Disgust; Intergroup Conflict, Models of; Political Psychology; Resentment

Further Readings

INTERGROUP CONFLICT, MODELS OF

Intergroup conflict typically originates in the rival interests of groups with distinct memberships. Beyond prejudice, and dislike that is exhibited in attitudes, its most harmful manifestations include social exclusion, violent clashes, street battles, ethnic conflict, and civil war that involve actions oriented against the out-group from in-group members. Relevant models all recognize that although groups might have their leadership or hierarchical organization, they are not unitary actors but rather are collections of individuals who, under social influence and control, have their free will and cognitive capacity to act in favor of their group's interest or form attitudes that stigmatize out-group members. Rivalry is symmetrical in the sense that a competitive action by in-group members hurts the out-group, but also a competitive action by the out-group hurts in-group members. While contributions to intergroup conflict in many competitive situations are intentional, the challenging task of models is to explain unintended and sub-optimal situations that should be avoided from the collective point of view. Among several theoretical models that have been proposed previously to understand and explain the causes, emergence, and dynamics of intergroup conflict, this entry focuses on key models and theories in sociology, social psychology, and economics.

Classical Models in Sociology

Karl Marx viewed historical development in terms of ongoing class conflict. Class conflict has its origins in differences in economic resources, and, in particular, in property ownership. Economic resources establish power and the exploitation of labor. The recognition of common interests leads to class consciousness that is a requirement of successful collective action and of revolution that breaks down existing class power.

In a different perspective, the seminal work of Georg Simmel emphasized the functional roles of intergroup conflict in group consciousness, in cohesion, in integration, and in solidarity within the group. Simmel argued that individuals have a limited choice in conflict situations, as the force to comply with a uniform group action is very strong, but he also underlined the importance of relational structures in conflict.

Drawing on Simmel's work, realistic conflict theory was oriented toward the functions of intergroup conflict. As Lewis Coser argued, the primary functions of conflict are to establish and maintain group identities and boundaries, and to support internal solidarity. The starting point of realistic conflict is intergroup competition. Groups compete with each other for certain scarce resources; it is the scarcity of these resources that makes intergroup conflict “realistic.” Groups try to obtain these resources because it is in their interest. In general, hostility toward the competing group is just a means for obtaining the scarce resources for the in-group.

Hubert M. Blalock handled actual intergroup competition for scarce resources at the macro level separately from individual actions. In his model, individuals are mobilized if they perceive intergroup competition or an out-group threat. This perception
Intergroup Conflict, Models of

Models of Lewis F. Richardson were based on statistical proportions of individuals with different war-moods (attitudes toward going to war) in the rival nations. He explained changes in the proportions by an underlying mechanism that is analogous to epidemics. Friendly and hostile attitudes spread irreversibly as a disease or fashion and result in transitions from one mood to another. When appropriate transition rules and starting parameters are assumed, the typical phases of war can be deduced. For certain values of the parameters, a balanced state of armament levels is derived. The other two typical scenarios are escalation of a runaway arms race and complete disarmament.

Theories and Models in Social Psychology

The positive association between positive attitudes toward in-group members and negative attitudes toward out-group members is emphasized by the theoretical concept of ethnocentrism. William Graham Sumner claimed that this association is universal and every group has a syndrome of ethnocentrism.

Muzafer Sherif, in his theory of conflict, focused on the emergence of negative out-group attitudes and intergroup hostility. In line with realistic conflict theory, he emphasized that competition between groups has a fundamental impact on the evolution of group structures, on negative out-group attitudes, and on the emergence of hostile behavior. Sherif also attempted to determine possible resolutions of conflict. He argued that occasional contact does not improve intergroup relations, but cooperative interdependencies and the establishment of superordinate, common goals do.

For the emergence of in-group favoritism and out-group hostility, it is not required that groups be organized or have an established set of norms. This is highlighted by the minimal group paradigm. Groups that are created on a cognitive basis and therefore are referred to as “minimal” are able to develop in-group favoritism even in the absence of social interaction. Examples from minimal group experiments are the groups created based on a preference for the paintings of Klee or Kandinsky, or based on the overestimation or underestimation of dots in a figure.

The minimal group paradigm stimulated the formation of social identity theory. The basic premise of social identity theory is that individuals strive to achieve or maintain a satisfactory image about themselves and that an important aspect of self-definition is group identification. Once people perceive group boundaries, cognitive processes trigger an in-group/out-group categorization. Accentuation of group boundaries results in polarization of the worthy in-group image and the unworthy out-group picture that together provide positive social identity. Individuals strive for and invest emotional energy to obtain social identity. This might ground their behavioral decision to participate in in-group/out-group hostility. Mostly, social identity is not taken into account purposefully, but it can indirectly explain why group members are mobilized into conflict.

As a follow-up to social identity theory, on one side, self-categorization theory emphasizes intergroup behavior in terms of underlying cognitive representations. Radical views express that individuals are “transformed into groups” via the process of self-categorization, meaning that this cognitive mechanism is sufficient for thinking in group terms. On the other side, critics of self-categorization theory maintain that social identity is obtained conditionally, only if the group is seen as an acting social unit, and these critics give more attention to the instrumental character of intergroup relations in which one group competes with another to achieve valued goals.

Rational Choice, Economics, and Game Theory

Ethnic competition theory emphasizes economic incentives in intergroup relations. Ethnic groups are considered as effective coalitions that are formed to extract material benefits from others or to defend possessions. This is also the source of ethnic competition and violence. The stronger the ethnic competition, the more severe the attitudes toward the competitor group are. Because employment possibilities are scarce and are the targets of ethnic competition, competition is intensified by desegregation of the labor market. Ethnic groups (and other groups) have fixed group boundaries, meaning that entry and exit are blocked. Ethnic networks therefore enjoy several advantages, like trust, cohesion, and easier establishment of collective action.
Individual decisions in the intergroup context are interdependent, because contributions to harmful intergroup competition have an effect on intergroup relations and consequently on the well-being of out-group members. This interdependence is captured by game-theoretic models of conflict. The simplest game-theoretic model is the two-person “prisoner’s dilemma.” In the prisoner’s dilemma, players have two options: cooperate or defect. Defection brings higher benefits irrespective of the action of the other party, but if both sides defect, payoffs are lower than if both sides cooperate. Endless civil war, unsettled disputes, and mutually harmful clashes are examples where both sides are punished and no benefits are realized. If groups were unitary entities and they could choose between cooperation and defection (e.g., the mobilization of group members to fight), mobilization would be their dominant strategy. This means that regardless of the action of the out-group, the in-group would gain more with mobilization. Following the dominant strategy by both sides would result in a suboptimal outcome.

A substantial element of intergroup relations is the free rider problem within the groups. Whether looking at civil war, violence between football supporters, or fights between urban gangs, participation in conflict involves high costs and risk for the participants. If they do not participate and instead let others “do the job,” individuals can still enjoy the benefits of intergroup competition (e.g., public happiness, pride, or social identity) that can be considered as group public goods. As individuals are able to free ride on the effort of others, in-group interdependence can be considered as a collective action problem. Recent models that take account of the free rider problem also recognize that individuals have contradictory interests considering intergroup and within-group relations. Conflict between the groups often originates from an establishment of in-group collective action. Successful mobilization is difficult to achieve; violent competitions are therefore less frequent than peaceful coexistence.

Which group is able to realize the benefits of intergroup competition depends on an intergroup comparison of the strength and efficiency of mobilization of group members. The appropriate model should represent the duality of within-group and intergroup interdependencies. A framework that allows for combined modeling is offered by the team games approach. In team games, competitive group relations are represented by a two-level game. At the intergroup level, groups as aggregates face an interdependent situation. Group action is determined as a consequence of individual actions within the group. At the group level, for groups with \( n \) people, strategic interdependence can be represented as an \( n \)-person game. The most relevant examples for models of intergroup conflict in the team games approach are the intergroup prisoner’s dilemma game, in which no participation is a dominant strategy equilibrium that is suboptimal in comparison to the outcome in which everyone participates; the intergroup public goods game, in which intergroup competition is based on the number of contributors and public good benefits are distributed in the group with more contributors; the intergroup chicken game; and the intergroup assurance game. Model extensions include structurally embedded team games that incorporate dyadic forms of social control in the team games model and that highlight the effect of social networks and segregation on intergroup conflict.

Károly Takács

See also Collective Action; Decision Theory, Philosophical Perspectives; Intergroup Conflict; Political Psychology

Further Readings

INTROSPECTION

You can learn about your own mind in the same way you learn about others’ minds—by observing behavior, by reading psychology books, by looking at readouts of brain activity. But it’s also generally
thought that you can also learn about your mind in a way no one else can, by introspection. But what exactly is introspection? No simple characterization is widely accepted. Introspection is a key concept in psychology and philosophy of mind, both as a process worth studying in its own right and as a method (of disputable validity) for learning about the mind. This entry will discuss the general features of introspection, several broad classes of approach to introspection, and the accuracy of introspective judgments.

Accounts of Introspection

General Features of Introspection

For a process to qualify as an “introspective” process as the term is ordinarily used by philosophers and most psychologists, it must be a process that yields judgments, knowledge, belief, or the like (for simplicity, this entry will just refer to judgments) about one’s own current or very recently past mental states or processes (for simplicity, this entry will just refer to states). Furthermore, it must be a process (unlike, say, inference or perception) that can only yield judgments about one’s own currently ongoing or recently past mental states and not other people’s mental states or mental states other than those (at most) a few moments past.

However, there are arguably non-introspective cognitive processes that can only yield judgments about one’s own currently ongoing or recently past mental states. For example, inferring that one likes hats from the fact that one feels a hat pressing on one’s forehead would not generally be regarded as an introspective process; nor is the process of conforming to a made-up or confabulated self-description simply to render that self-description true; nor is any automatic, subpersonal interaction of different physiological or functional regions of the mind. Thus, there are three further conditions that are sometimes treated as necessary for a process to qualify as introspective: the directness condition, which requires that introspection yield judgments (knowledge, belief, or the like) by relatively direct means (and not, e.g., by inference from how things stand outside the mind); the detection condition, which requires that introspection involve some sort of detection of, or attunement to, a preexisting mental state or event (and not, e.g., a mental state that is brought into existence in the course of introspecting); and the effort condition, which requires that introspection not be constant, effortless, automatic, or subpersonal, but rather requires some sort of special reflection by the individual that is different from the ordinary un-self-reflective flow of thought and action in daily life. Scholars differ in the weight they put on these conditions, and because of this, an account of self-knowledge that violates one of these latter three conditions may be seen by some researchers but not others as an account of “introspection.”

Assertions of First- and Third-Person Parity

It is sometimes asserted that our only methods for knowing our own minds, or some aspect of our own minds, are the methods by which we know others’ minds. If so, then there is no distinctive introspective process or no such process that offers access to the sorts of mental states in question. No prominent psychologist or philosopher has embraced the position in its starkest and most universal form, although Daryl Bem’s self-perception theory comes close. According to self-perception theory, we learn about our own minds almost exclusively by observing our behavior, with little input from internal cues. For example, we infer what our attitudes are by noticing what we endorse and reject and how much money it requires to entice us to do something, just as we learn about the attitudes of other people. Richard Nisbett and Timothy Wilson have argued that we have no introspective access to the processes underlying our decisions and attitudes (although we do have uniquely first-person access to mental “content” such as current sensations and evaluations) and that we arrive at judgments about such things as our motives and reasons on pretty much the same basis as, and with no more accuracy than, we arrive at judgments about other people’s motives and reasons. However, Bem’s view is out of favor and Nisbett and Wilson’s remains contentious.

Personality traits (such as extraversion) and non-conscious subpersonal processes (such as early visual processing) are good candidates for mental states to which people have no special introspective access. If we know better than do others whether we are extraverted or not, it is because we have seen more of our own behavior and perhaps because we can introspect such things as incipient emotional reactions to imagined events, from which introspections
we can reach non-introspective inferential judgments about personality.

**Self-Monitoring Accounts of Introspection**

Introspection is often characterized as a type of mental monitoring, perhaps by a functionally distinct cognitive system or group of systems. Two important recent accounts are the account of Shaun Nichols and Stephen Stich, on the one hand, and the account of Alvin Goldman on the other.

The Nichols and Stich account of self-knowledge appeals to two distinct mechanisms. The first is a monitoring mechanism that simply converts internally stored representations with the content $P$ (e.g., the representation with the content *It will rain on Thursday*) into internally stored representational beliefs with the content $I A$ that $P$, where $A$ here refers to the attitude one takes toward that content—for example, belief or desire (I may believe that *it will rain on Thursday* or, alternatively, I may desire that it will). The second mechanism is the (non-introspective) theory of mind mechanism that we typically use in attributing mental states to other people but which can also be turned on ourselves. For example, if I can discern from your behavior that you enjoy sherry, I can also use the theory of mind mechanism to discern that fact from my own behavior. Nichols and Stich support the idea that there are two distinct mechanisms here by appealing to a double dissociation—that is, to cases in which it appears that one mechanism but not the other breaks down. They argue that in certain types of schizophrenia, the monitoring mechanism breaks down while the theory of mind mechanism remains intact and that in autism the theory of mind mechanism breaks down while the monitoring mechanism remains intact.

Goldman criticizes the Nichols and Stich account for failing to describe how the monitoring mechanism detects the attitude type of the detected representation (i.e., whether the attitude toward that representation is one of belief, desire, etc.) and for leaving unclear how people can discern the strength or intensity of their attitudes. His positive account starts from the idea that we can direct an attentional process toward discerning the general types of our mental states (e.g., belief, happiness, bodily sensation) and some of the properties of those mental states (e.g., intensity for belief, a variety of finely delineated categories for bodily sensation). However, Goldman argues that specific contents of attitudes like belief are too manifold for preexisting classification categories to exist for all of them. Rather, he suggests, we represent the specific contents of such mental states by “redeploying” the content of the represented mental state into the representation of that state—simply copying the representational content of the lower order state into the higher order state. (In this last respect, Goldman’s account resembles the Nichols and Stich account.)

**Introspection Without Self-Monitoring?**

A number of philosophers have developed accounts of introspection that jettison the intuitive idea that introspection is a type of self-monitoring. Sydney Shoemaker, for example, has developed an account of introspection on which it is just partly constitutive of believing something that you are disposed to judge that you believe it—no special act of self-monitoring required. Other philosophers have endorsed what they call the “transparency” of self-knowledge—the idea that, when asked about one’s own mental states (e.g., “Do you think there will be a Third World War?”) one tends to reflect not on one’s own mind but on the world outside one’s mind (e.g., on whether political conditions and human nature are such that a Third World War is likely and consequently saying “yes” or “no” in answer to the question about what one thinks). Whereas some transparency theorists deny that the process of looking outward or thinking about the external world to answer questions about one’s mental states is, strictly speaking, an “introspective” process, others such as Fred Dretske have characterized such accounts as introspective—thus manifesting dissent about whether the detection and directness conditions previously described are genuinely necessary conditions on an “introspective” process.

**The Accuracy of Introspective Judgments**

Although there is some dissent on the issue, a majority of philosophers have long held that knowledge of one’s own mental states—or at least some important subclass of them, such as one’s currently ongoing conscious states—is highly trustworthy, perhaps even (as René Descartes suggested in the
17th century) infallible or indubitable. Psychologists, in contrast, have a history of skepticism about the accuracy of introspective reports, dating back to the early introspective psychologists of the late 19th and early 20th centuries, such as Wilhelm Wundt and E. B. Titchener, who repeatedly cautioned about the difficulty of introspective observation and developed training techniques intended to improve the accuracy of introspective reports.

Contemporary psychologists criticizing the accuracy of introspection have often cited the work of Nisbett and Wilson mentioned earlier, which seems to suggest that people’s access to their own motives and reasons is neither especially better nor different in kind from the access other people have. Whether this shows that introspection is inaccurate, or instead simply that we do not or cannot strictly speaking introspect our motives and reasons, is unclear; in any case, if the Nisbett and Wilson view is roughly correct, whatever process it is that people engage in when they report on their motives and reasons is not an especially accurate one.

Often cited as suggesting the accuracy of a certain sort of introspective report—although again the question arises of whether what is involved is “introspection” strictly speaking—is the work of K. Anders Ericsson and Herbert Simon on “think-aloud protocols” and retrospective verbalizations of thought processes. Ericsson and Simon review evidence suggesting that people often engage in the cognitive processes they say they do, when asked to give simultaneous or immediately retrospective description of those processes, and thus that verbal reports about such processes may be valuable data.

Finally, in the interdisciplinary field of “consciousness studies,” considerable attention has recently been devoted to the accuracy or inaccuracy of introspective reports about the stream of conscious experience, including imagery, sensory experience, and emotional experience—without, so far, any clear resolution.

Eric Schwitzgebel

See also Access Consciousness; Inner Speech; Self-Consciousness; Self-Knowledge; Unconscious Emotions, Psychological Perspectives

Further Readings


ENCYCLOPEDIA of the MIND
Editorial Board

Editor-in-Chief

Harold Pashler
University of California, San Diego

Associate Editor

Emily Goldknopf
University of California, Los Angeles

Editorial Board

Tim Crane
University of Cambridge

Fernanda Ferreira
University of Edinburgh

Marcel Kinsbourne
New School for Social Research

Richard Zemel
University of Toronto

Editorial Assistant

Ruth Hsiao
University of Illinois College of Medicine
Contents

Volume 2
List of Entries  vii
Reader’s Guide  xi

Entries

<table>
<thead>
<tr>
<th>J</th>
<th>441</th>
<th>R</th>
<th>627</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>447</td>
<td>S</td>
<td>663</td>
</tr>
<tr>
<td>L</td>
<td>455</td>
<td>T</td>
<td>731</td>
</tr>
<tr>
<td>M</td>
<td>477</td>
<td>U</td>
<td>749</td>
</tr>
<tr>
<td>N</td>
<td>553</td>
<td>V</td>
<td>757</td>
</tr>
<tr>
<td>O</td>
<td>569</td>
<td>W</td>
<td>775</td>
</tr>
<tr>
<td>P</td>
<td>577</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Index  797
List of Entries

Access Consciousness
Action and Bodily Movement
Action Slips
Addiction
Affective Forecasting
Aging, Memory, and Information Processing Speed
Allais Paradox
Amnesia
Analogical Mapping and Reasoning
Analogical Reasoning, Models of Development
Anchoring
Anesthesia and Awareness
Anomalous Monism
Anosognosia
Anti-Individualism About Cognition
Anxiety Disorders
Aphasia
Apraxia
Argument Mapping
Atomism About Concepts
Attention, Evolutionary Perspectives
Attention, Neuroimaging Studies of
Attention, Resource Models
Attention and Action
Attention and Consciousness
Attention and Emotion
Attention and Emotions, Computational Perspectives
Attentional Blink Effect
Attitude Change
Attitudes and Behavior
Attraction
Attribution Theory
Audition, Neural Basis
Auditory Masking
Autism
Automatic Behavior
Automaticity
Autoscopic Phenomena
Availability Heuristic
Behavioral Therapy
Behaviorism
Belief and Judgment
Bilingual Language Processing
Bilingualism, Cognitive Benefits of
Black English Vernacular (Ebonics)
Blindsight
Borderline Personality Disorder
Capgras Delusion
Case-Based Reasoning, Computational Perspectives
Categorization, Neural Basis
Categorization, Psychological Perspectives
Category Learning, Computational Perspectives
Causal Theories of Intentionality
Causal Theories of Memory
Change Blindness
Character and Personality, Philosophical Perspectives
Classical Theory of Concepts
Cognitive Dissonance
Cohort Model of Auditory Word Recognition
Collective Action
Common Coding
Compound Words, Processing of
Computational Models of Emotion
Concepts, Comparative Perspectives
Concepts, Development of
Concepts, Philosophical Issues
Concepts and Language
Conceptual Combination
Conduction Aphasia
Confabulation
Conscious Thinking
Consciousness, Comparative Perspectives
Consciousness and Embodiment
Consciousness and the Unconscious
Conversation and Dialogue
Debiasing
Deception, Linguistic Cues to
Decision Improvement Technologies
Decision Making, Neural Underpinnings
Decision Making and Reward, Computational Perspectives
Decision Theory, Philosophical Perspectives
Declarative/Procedural Model of Language
Deductive Reasoning
Delusions
Depth Perception
Descriptions
Descriptive Thought
Desirable Difficulties Perspective on Learning
Desire
Dictator Game
Discourse Processing, Models of
Discrimination Learning, Training Methods
Disfluencies: Comprehension Processes
Disgust
Disjunctive Theory of Perception
Dissent, Effects on Group Decisions
Distributed Cognition
Divided Attention and Memory
Dutch Book Arguments
Dyslexia, Acquired
Dyslexia, Developmental
Dyslexia, Phonological Processing in
Electrophysiological Studies of Mind
Eliminative Materialism
Embarrassment
Emergence
Emotion, Cerebral Lateralization
Emotion, Cultural Perspectives
Emotion, Psychophysiology of
Emotion, Structural Approaches
Emotion and Moral Judgment
Emotion and Psychopathology
Emotion and Working Memory
Emotion Regulation
Emotional Recognition, Neuropsychology of
Emotions and Consciousness
Endowment Effect
Envy
Event Memory, Development
Exercise and the Brain
Experimental Philosophy
Explanation of Action
Explanatory Gap
Extended Mind
Eye Movements During Reading
Eyewitness Memory
Face Perception
Face Recognition in Humans and Computers
Facial Expressions, Computational Perspectives
Facial Expressions, Emotional
Flynn Effect: Rising Intelligence Scores
Folk Psychology
Freedom of Action
Fregoli Delusion
Frequency Effects in Word Recognition
Gender Differences in Language and Language Use
Genes and Linguistic Tone
Gesture and Language Processing
Group Decision Making
Guilt
Happiness
Hearing
Hearing, Philosophical Perspectives
Heritability
Heritage Language and Second Language Learning
Human Classification Learning
Hypochondria
Idealism
Implicit Memory
Inattentional Blindness
Indexical Thought
Inferences in Language Comprehension
Inhibition of Return
Innateness and Parameter Setting
Inner Speech
Intelligence, Neural Basis
Intelligence and Working Memory
Intension and Extension
Intentionality of Bodily Sensation
Intentionality of Emotion
Intergroup Conflict
Intergroup Conflict, Models of
Introspection
Jealousy
Joint or Collective Intention
Know-How, Philosophical Perspectives
Knowledge Acquisition in Development
Knowledge by Acquaintance
Language Development
Language Development, Overregulation in
Language Production, Agreement in
Language Production, Incremental Processing in
Layered Control Architectures
Learning Styles
Legal Reasoning, Psychological Perspectives
Lie Detection
Love

Machine Speech Recognition
McCollough Effect
Memory, Interference With
Memory, Neural Basis
Memory and Knowledge
Memory Recall, Dynamics
Mental Action
Mental Causation
Mental Effort
Meta-Analysis
Metacognition and Education
Metaphor
Microgenesis of Consciousness
Mind-Body Problem
Mirror Neurons
Mnemonic Strategies
Modeling Causal Learning
Moral Development
Motivated Thinking
Motor Learning, Practical Aspects
Motor System, Development of
Multimodal Conversational Systems
Multinomial Modeling
Multiple Intelligences Theory
Multitasking and Human Performance
Music and the Evolution of Language
Music Perception

Naïve Realism
Narcissistic Personality Disorder
Natural Action Selection, Modeling
Natural Language Generation
Neural Correlates of Consciousness
Neurodynamics of Visual Search
Neuroeconomics

Object-Dependent Thought
Objects of Memory
Optic Flow

Perceptual Consciousness and Attention
Perceptual Constancy
Personal Identity
Personal Identity, Development of
Personal Identity, Development of Personality: Individual Versus Situation Debate
Perspective Taking in Language Processing
Persuasion
Phenomenology of Action
Philosophy of Action
Physicalism
Placebo Effect
Planning in Language Production
Political Psychology
Preconscious Free Will
Prediction, Clinical Versus Actuarial
Production of Language
Prosody in Production
Psychological Refractory Period

Rationality of Emotion
Reaction Time
Realism and Instrumentalism
Reductive Physicalism
Regret
Rehearsal and Memory
Reinforcement Learning, Psychological Perspectives
Relationships, Development of
Religion and Psychiatry
Representational Theory of Mind
Representations, Development of
Representativeness Heuristic
Resentment
Retrieval Practice (Testing) Effect

Schizophrenia
Scientific Reasoning
Self, Development of
Self-Consciousness
Self-Knowledge
Semantic Dementia
Semantic Memory
Semantic Memory, Computational Perspectives
Sequential Memory, Computational Perspectives
Serial Order Memory, Computational Perspectives
Similarity
Skill Learning, Enhancement of
Sleep and Dreams
Smell, Philosophical Perspectives
<table>
<thead>
<tr>
<th>Social Cognition</th>
<th>Unconscious Emotions, Psychological Perspectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Loafing</td>
<td>Unconscious Perception</td>
</tr>
<tr>
<td>Spacing Effect</td>
<td></td>
</tr>
<tr>
<td>Spacing Effect, Practical Applications</td>
<td>Visual Imagery</td>
</tr>
<tr>
<td>Spatial Cognition, Development of Speech Perception</td>
<td>Visual Masking</td>
</tr>
<tr>
<td>Stereopsis</td>
<td>Visual Search</td>
</tr>
<tr>
<td>Stroop Effect</td>
<td>Visual Working Memory</td>
</tr>
<tr>
<td>Subliminal Perception</td>
<td>Visuospatial Reasoning</td>
</tr>
<tr>
<td>Synesthesia</td>
<td>Voluntary Action, Illusion of</td>
</tr>
<tr>
<td>Syntactic Production, Agreement in Taste, Philosophical Perspectives</td>
<td>Williams Syndrome</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Wisdom of Crowds Effect</td>
</tr>
<tr>
<td></td>
<td>Word Learning</td>
</tr>
<tr>
<td></td>
<td>Word Recognition, Auditory</td>
</tr>
<tr>
<td></td>
<td>Word Recognition, Visual</td>
</tr>
<tr>
<td></td>
<td>Working Memory</td>
</tr>
<tr>
<td></td>
<td>Working Memory, Evolution of</td>
</tr>
<tr>
<td></td>
<td>Working Memory in Language Processing</td>
</tr>
</tbody>
</table>
Reader’s Guide

Action and Motor Control

Computational Perspectives
Layered Control Architectures
Natural Action Selection, Modeling

Development
Motor System, Development of

Disorders and Pathology
Apraxia

Neural Basis
Apraxia
Common Coding
Desire
Mirror Neurons
Preconscious Free Will

Philosophical Perspectives
Action and Bodily Movement
Collective Action
Desire
Explanation of Action
Freedom of Action
Joint or Collective Intention
Know-How, Philosophical Perspectives
Mental Action
Phenomenology of Action
Philosophy of Action
Preconscious Free Will
Teleology

Psychological Research
Action Slips
Attention and Action
Common Coding
Motor System, Development of
Multitasking and Human Performance
Optic Flow

Preconscious Free Will
Psychological Refractory Period
Voluntary Action, Illusion of

Attention

Computational Perspectives
Attention and Emotions, Computational Perspectives

Evolutionary Perspectives
Attention, Evolutionary Perspectives

Neural Basis
Attention, Evolutionary Perspectives
Attention, Neuroimaging Studies of Inhibition of Return
Mental Effort
Neurodynamics of Visual Search

Philosophical Perspectives
Attention and Consciousness

Psychological Research
Attention, Resource Models
Attention and Action
Attention and Emotion
Attentional Blink Effect
Automaticity
Change Blindness
Divided Attention and Memory
Inattentional Blindness
Inhibition of Return
Mental Effort
Multitasking and Human Performance
Neurodynamics of Visual Search
Perceptual Consciousness and Attention
Psychological Refractory Period
Stroop Effect
Visual Search
Concepts and Categories
Computational Perspectives
Category Learning, Computational Perspectives
Semantic Memory, Computational Perspectives

Development
Concepts, Development of
Representations, Development of
Self, Development of

Disorders and Pathology
Semantic Dementia

Evolutionary Perspectives
Concepts, Comparative Perspectives

Neural Basis
Categorization, Neural Basis

Philosophical Perspectives
Atomism About Concepts
Causal Theories of Intentionality
Classical Theory of Concepts
Concepts, Philosophical Issues
Concepts and Language
Microgenesis of Consciousness
Representational Theory of Mind

Psychological Research
Categorization, Psychological Perspectives
Concepts, Comparative Perspectives
Conceptual Combination
Self, Development of
Semantic Memory
Similarity

Consciousness
Disorders and Pathology
Anosognosia
Automatic Behavior
Autoscopic Phenomena

Evolutionary Perspectives
Consciousness, Comparative Perspectives

Neural Basis
Anesthesia and Awareness
Anosognosia
Microgenesis of Consciousness

Neural Correlates of Consciousness
Preconscious Free Will
Self-Consciousness
Sleep and Dreams

Philosophical Perspectives
Access Consciousness
Attention and Consciousness
Conscious Thinking
Consciousness and Embodiment
Consciousness and the Unconscious
Emotions and Consciousness
Explanatory Gap
Introspection
Microgenesis of Consciousness
Realism and Instrumentalism
Self-Consciousness
Self-Knowledge

Psychological Research
Anesthesia and Awareness
Attention and Consciousness
Consciousness and the Unconscious
Metacognition and Education
Perceptual Consciousness and Attention
Self-Consciousness
Self-Knowledge
Subliminal Perception
Unconscious Emotions, Psychological Perspectives
Unconscious Perception
Voluntary Action, Illusion of

Decision Making
Computational Perspectives
Decision Making and Reward, Computational Perspectives
Natural Action Selection, Modeling

Neural Basis
Decision Making, Neural Underpinnings
Dictator Game
Neuroeconomics

Philosophical Perspectives
Decision Theory, Philosophical Perspectives
Dutch Book Arguments

Practical Applications
Debiasing
Decision Improvement Technologies
Psychological Research
Affective Forecasting
Allais Paradox
Cognitive Dissonance
Debiasing
Dictator Game
Dissent, Effects on Group Decisions
Endowment Effect
Group Decision Making
Wisdom of Crowds Effect

Disorders and Pathology
Addiction
Amnesia
Anosognosia
Anxiety Disorders
Aphasia
Apraxia
Autism
Automatic Behavior
Autoscopic Phenomena
Behavioral Therapy
Borderline Personality Disorder
Capgras Delusion
Conduction Aphasia
Confabulation
Delusions
Dyslexia, Acquired
Dyslexia, Developmental
Dyslexia, Phonological Processing in
Emotion and Psychopathology
Fregoli Delusion
Hypochondria
Narcissistic Personality Disorder
Religion and Psychiatry
Schizophrenia
Semantic Dementia
Williams Syndrome

Emotion
Computational Perspectives
Attention and Emotions, Computational Perspectives
Computational Models of Emotion
Facial Expressions, Computational Perspectives

Cultural Perspectives
Emotion, Cultural Perspectives

Disorders and Pathology
Anxiety Disorders
Religion and Psychiatry

Evolutionary Perspectives
Consciousness, Comparative Perspectives

Neural Basis
Desire
Emotion, Cerebral Lateralization
Emotion, Psychophysiology of
Emotion and Psychopathology
Emotional Recognition, Neuropsychology of

Philosophical Perspectives
Desire
Emotion and Moral Judgment
Emotions and Consciousness
Intentionality of Emotion
Rationality of Emotion

Psychological Research
Affective Forecasting
Anxiety Disorders
Attention and Emotion
Cognitive Dissonance
Disgust
Embarrassment
Emotion, Psychophysiology of
Emotion, Structural Approaches
Emotion and Moral Judgment
Emotion and Working Memory
Emotion Regulation
Envy
Facial Expressions, Emotional
Guilt
Happiness
Jealousy
Love
Regret
Religion and Psychiatry
Resentment
Unconscious Emotions, Psychological Perspectives

Epistemology
Philosophical Perspectives
Anti-Individualism About Cognition
Belief and Judgment
Causal Theories of Memory
Conscious Thinking
Consciousness and the Unconscious
Deductive Reasoning
Descriptive Thought
Emotion and Moral Judgment
Experimental Philosophy
Explanatory Gap
Introspection
Know-How, Philosophical Perspectives
Knowledge by Acquaintance
Memory and Knowledge
Self-Knowledge

Psychological Research
Self-Knowledge

Genetics
Heritability

Disorders and Pathology
Addiction
Schizophrenia
Williams Syndrome

Evolutionary Perspectives
Genes and Linguistic Tone

Groups
Cultural Perspectives
Distributed Cognition

Philosophical Perspectives
Collective Action
Joint or Collective Intention

Psychological Research
Dissent, Effects on Group Decisions
Distributed Cognition
Group Decision Making
Intergroup Conflict
Intergroup Conflict, Models of

Intelligence
Neural Basis
Intelligence, Neural Basis

Psychological Research
Flynn Effect: Rising Intelligence Scores
Heritability
Intelligence and Working Memory
Multiple Intelligences Theory

Language and Communication
Computational Perspectives
Machine Speech Recognition
Multimodal Conversational Systems
Natural Language Generation

Cultural Perspectives
Bilingual Language Processing
Bilingualism, Cognitive Benefits of
Black English Vernacular (Ebonics)
Gender Differences in Language and Language Use
Gesture and Language Processing
Heritage Language and Second Language Learning

Development
Dyslexia, Developmental
Dyslexia, Phonological Processing in
Heritage Language and Second Language Learning
Innateness and Parameter Setting
Language Development
Language Development, Overregulation in Word Learning

Disorders and Pathology
Aphasia
Conduction Aphasia
Dyslexia, Acquired
Dyslexia, Developmental
Dyslexia, Phonological Processing in Semantic Dementia

Evolutionary Perspectives
Genes and Linguistic Tone
Music and the Evolution of Language

Neural Basis
Aphasia
Conduction Aphasia
Philosophical Perspectives
Descriptions
Descriptive Thought
Indexical Thought
Intension and Extension
Metaphor
Object-Dependent Thought
Practical Applications
Deception, Linguistic Cues to
Lie Detection
Psychological Research
Bilingual Language Processing
Bilingualism, Cognitive Benefits of
Cohort Model of Auditory Word Recognition
Compound Words, Processing of
Concepts and Language
Conceptual Combination
Conversation and Dialogue
Declarative/Procedural Model of Language
Discourse Processing, Models of
Disfluencies: Comprehension Processes
Eye Movements During Reading
Frequency Effects in Word Recognition
Gesture and Language Processing
Inferences in Language Comprehension
Inner Speech
Language Development
Language Production, Agreement in
Language Production, Incremental
Processing in
Lie Detection
Metaphor
Perspective Taking in Language Processing
Planning in Language Production
Production of Language
Prosody in Production
Speech Perception
Syntactic Production, Agreement in
Word Learning
Word Recognition, Auditory
Word Recognition, Visual
Working Memory in Language Processing

Learning
Computational Perspectives
Modeling Causal Learning

Development
Knowledge Acquisition in Development
Word Learning

Neural Basis
Desire
Exercise and the Brain
Human Classification Learning

Practical Applications
Behavioral Therapy
Desirable Difficulties Perspective on Learning
Discrimination Learning, Training Methods
Exercise and the Brain
Learning Styles
Metacognition and Education
Motor Learning, Practical Aspects
Placebo Effect
Retrieval Practice (Testing) Effect
Skill Learning, Enhancement of
Spacing Effect, Practical Applications

Psychological Research
Behavioral Therapy
Behaviorism
Desirable Difficulties Perspective on Learning
Desire
Human Classification Learning
Learning Styles
Metacognition and Education
Placebo Effect
Reinforcement Learning, Psychological Perspectives
Retrieval Practice (Testing) Effect
Skill Learning, Enhancement of
Spacing Effect
Word Learning

Memory
Computational Perspectives
Semantic Memory, Computational Perspectives
Sequential Memory, Computational Perspectives
Serial Order Memory, Computational Perspectives

Development
Aging, Memory, and Information Processing Speed
Event Memory, Development
Disorders and Pathology
Amnesia
Confabulation
Semantic Dementia

Evolutionary Perspectives
Working Memory, Evolution of

Neural Basis
Amnesia
Event Memory, Development
Implicit Memory
Memory, Neural Basis

Philosophical Perspectives
Causal Theories of Memory
Memory and Knowledge
Objects of Memory

Practical Applications
Exercise and the Brain
Retrieval Practice (Testing) Effect
Mnemonic Strategies
Spacing Effect, Practical Applications

Psychological Research
Aging, Memory, and Information Processing
Speed
Divided Attention and Memory
Emotion and Working Memory
Event Memory, Development
Exercise and the Brain
Eyewitness Memory
Implicit Memory
Intelligence and Working Memory
Memory, Interference With
Memory Recall, Dynamics
Rehearsal and Memory
Retrieval Practice (Testing) Effect
Semantic Memory
Spacing Effect
Visual Working Memory
Working Memory
Working Memory in Language Processing

Metaphysics
Philosophical Perspectives
Anomalous Monism
Anti-Individualism About Cognition

Causal Theories of Intentionality
Conscious Thinking
Consciousness and Embodiment
Eliminative Materialism
Emergence
Extended Mind
Freedom of Action
Idealism
Intentionality of Bodily Sensation
Intentionality of Emotion
Mental Action
Mental Causation
Naïve Realism
Object-Dependent Thought
Personal Identity
Philosophy of Action
Physicalism
Realism and Instrumentalism
Reductive Physicalism
Smell, Philosophical Perspectives
Taste, Philosophical Perspectives
Teleology
Theory of Appearing

Methodology
Neural Basis
Electrophysiological Studies of Mind

Philosophical Perspectives
Experimental Philosophy

Practical Applications
Prediction, Clinical Versus Actuarial

Psychological Research
Heritability
Meta-Analysis
Multinomial Modeling
Reaction Time

Mind-Body Problem
Philosophical Perspectives
Action and Bodily Movement
Anomalous Monism
Consciousness and Embodiment
Eliminative Materialism
Emergence
Explanatory Gap
Idealism
Know-How, Philosophical Perspectives
Mental Causation
Mind-Body Problem
Physicalism
Realism and Instrumentalism
Reductive Physicalism
Representational Theory of Mind
Teleology

**Morality/Ethics**

*Development*
Moral Development

**Philosophical Perspectives**
Character and Personality, Philosophical Perspectives
Emotion and Moral Judgment
Joint or Collective Intention
Philosophy of Action

**Psychological Research**
Emotion and Moral Judgment
Guilt
Moral Development
Political Psychology

**Perception**

*Computational Perspectives*
Face Recognition in Humans and Computers
Facial Expressions, Computational Perspectives

**Neural Basis**
Audition, Neural Basis
Blindsight
Common Coding
Emotional Recognition, Neuropsychology of Mirror Neurons
Synesthesia
Visual Imagery

**Philosophical Perspectives**
Disjunctive Theory of Perception
Hearing, Philosophical Perspectives
Naïve Realism
Smell, Philosophical Perspectives
Taste, Philosophical Perspectives

*Psychological Research*
Auditory Masking
Common Coding
Depth Perception
Face Perception
Facial Expressions, Emotional
Hearing
McCollough Effect
Music Perception
Optic Flow
Perceptual Consciousness and Attention
Perceptual Constancy
Speech Perception
Stereopsis
Subliminal Perception
Synesthesia
Time Perception
Unconscious Perception
Visual Imagery
Visual Masking
Visual Search
Visuospatial Reasoning
Word Recognition, Auditory
Word Recognition, Visual

**Personality**

*Development*
Personal Identity, Development of

**Disorders and Pathology**
Borderline Personality Disorder
Narcissistic Personality Disorder

**Philosophical Perspectives**
Character and Personality, Philosophical Perspectives

*Psychological Research*
Attitude Change
Borderline Personality Disorder
Personal Identity, Development of Personality: Individual Versus Situation Debate
Political Psychology

**The Social Mind**

*Development*
Relationships, Development of
Disorders and Pathology
Capgras Delusion

Philosophical Perspectives
Folk Psychology

Psychological Research
Attitude Change
Attitudes and Behavior
Attraction
Attribution Theory
Cognitive Dissonance
Group Decision Making
Intergroup Conflict
Intergroup Conflict, Models of
Love
Personality: Individual Versus Situation
Debate
Persuasion
Political Psychology
Social Cognition
Social Loafing
Wisdom of Crowds Effect

Thinking and Reasoning
Computational Perspectives
Analogical Reasoning, Models of Development
Case-Based Reasoning, Computational Perspectives

Cultural Perspectives
Distributed Cognition

Development
Analogical Reasoning, Models of Development
Knowledge Acquisition in Development
Spatial Cognition, Development of

Philosophical Perspectives
Deductive Reasoning

Practical Applications
Argument Mapping
Case-Based Reasoning, Computational Perspectives

Psychological Research
Analogical Mapping and Reasoning
Anchoring
Availability Heuristic
Distributed Cognition
Legal Reasoning, Psychological Perspectives
Motivated Thinking
Representativeness Heuristic
Scientific Reasoning
Spatial Cognition, Development of Thinking
Two System Models of Reasoning
Visual Imagery
Visuospatial Reasoning
JEALOUSY

Jealousy is a subjectively unpleasant emotion that occurs when one perceives that an important aspect of one’s relationship with another, or the relationship itself, is being threatened by a third person. Hence, jealousy requires the involvement of three individuals (a “love triangle”): the self, the loved one (partner), and the rival. Jealousy can occur in romantic relationships over acts such as infidelity. It also occurs in other forms of relationships such as when children feel upset over a parent showering attention on a new sibling, or when a person feels distress due to being excluded by friends who are socializing together. The proposed function of jealousy is to motivate behaviors that will protect or maintain the relationship between the self and the partner and reduce the threatening bond between the partner and rival.

Although jealousy may lead to desirable outcomes such as redirecting a loved one’s attention to the self and reestablishing bonds, it can also have serious negative consequences. For example, jealousy is often implicated as a cause of spousal abuse and is the third or fourth most common motive in non-accidental homicides across cultures. This entry discusses theories of jealousy, including conceptual debates about its origin and definition, and presents research on the development of jealousy and individual differences in jealousy.

Theories

Theorists agree that jealousy involves unpleasant feelings, but there is no unanimity on the precise nature of the distress. One possibility is that the feelings commonly referred to as jealousy may be a blend of other emotions such as anger, fear, and sadness. There are two routes by which this could occur: (1) People may simultaneously experience several emotions during a jealous episode, or (2) they may experience a series of different emotions over the course of a single jealousy episode. In the latter case, the emotion felt at any given moment would depend on the aspect of the situation on which the person focused. For example, contemplating the loss of the relationship might elicit sadness, whereas thinking about the partner’s dishonesty might elicit anger. It is also possible that jealousy is a unique emotional state that produces its own distinct feelings and behaviors that differ from other emotions such as fear and anger.

Development

Signs of jealousy have been found in children as young as 6 months when their mothers directed attention to what appeared to be another baby. This suggests that at least some primitive forms of jealousy can be elicited without complex thoughts. However, with cognitive development, the triggers for jealousy become more sophisticated. For example, one study found that 4-year-old children showed
more jealousy when their mothers interacted with a similar-aged peer than when she interacted with an infant. Jealousy in younger infants was not affected by the rival's age. Thus, it appears that over the course of development, an individual's appraisals of the nature and meaning of the interactions between the rival and the loved one become increasingly important in the elicitation of jealousy.

Social-cognitive theorists have focused on two general factors that cause a loved one’s involvement with another to be particularly upsetting: (a) when it reduces benefits obtained from the primary relationship, and (b) when it threatens some aspect of a person’s self-concept or self-esteem. People ponder the meaning and ramifications of their loved one’s relationship to the rival—“Will my partner stop giving me time and attention?” and “What does this mean about me? Am I unattractive or unlovable?” Conclusions that people draw from such questions affect the intensity and nature of their jealousy.

Individual Differences

Attachment Styles

According to attachment theory, people’s experiences, beginning in infancy, lead them to form mental models of relationships that include beliefs about others and the self. People who have a secure attachment style readily trust others and are comfortable with intimacy. Research suggests that differences in attachment style may play an important role in jealous reactions. For example, one study found that securely attached individuals reported that a past jealousy experience brought them closer to their partners—an effect not experienced by individuals with insecure attachment styles.

Gender

There is controversy over whether men and women are jealous over different things. The jealousy as a specific module view hypothesizes that women should feel more jealous over emotional betrayal and men over sexual betrayal because the two genders face different reproductive threats. (The basic tenet of modern evolutionary theory is that we inherited our psychological and/or physical traits from the ancestral people who reproduced the most, i.e., had higher inclusive fitness.) Because fertilization occurs internally within women, men could never know for certain whether an offspring was biologically their own. Therefore, men should be particularly concerned about a mate’s sexual infidelity because it could lead to the man expending valuable resources (food) on offspring that were not genetically his own, which would be costly to his inclusive fitness. Ancestral woman faced a different threat; she needed to ensure that her mate did not give his resources to other women and their children, which could decrease the likelihood of the woman’s own children surviving. Thus, present-day women should be particularly jealous over emotional infidelity. Inherent in this is the assumption that a man’s emotional involvement is a proxy for his spending resources on another.

This hypothesis drew apparent support from early work that found when forced to predict whether a partner’s sexual or emotional infidelity would be more upsetting, more women than men picked emotional infidelity. However, several lines of new research with other measures and with participants who have actually experienced a loved one’s betrayal have not found consistent gender differences in reactions to sexual and emotional infidelity. For example, one study found that men and women, regardless of sexual orientation, focused more on the emotional aspects of their partner’s actual betrayal relative to the sexual aspects.

This begs the question of why men and women have similar jealousy reactions. One possibility is there may have been no need for sexually dimorphic jealousy mechanisms—a more general jealousy process may have addressed the inclusive fitness risks faced by either gender. Flirting behaviors (increased eye contact, smiling) usually occur well before having sex or falling in love. Because the same behaviors can signal the beginnings of emotional interest, sexual interest, or both, attention to these common early warning signs could enable both men and women to prevent their partners from engaging in either form of infidelity, without the need for sexual dimorphic mechanisms.

Christine R. Harris and Ryan S. Darby

See also Emotion Regulation; Envy; Rationality of Emotion; Relationships, Development of
JOINT OR COLLECTIVE INTENTION

Among those attitudes of individual human beings that have attracted the attention of philosophers are their intentions. The branch of philosophy that deals with such personal intentions is generally referred to as action theory. This may itself be considered a part of the philosophy of mind. In everyday speech, people talk not only of what "I" intend but also of what "we" intend. This suggests that there are not only personal intentions, that is, intentions of an individual person, but also joint or collective or shared intentions, that is, intentions of two or more individual people. The present discussion offers an introduction to contemporary philosophical discussions of joint intention, discussions that have been of interest to social scientists and others in a variety of fields, including social and developmental psychology and cognitive science.

Joint Intentions With Regard to the Future Versus Joint Intentions in Acting

Theorists of personal intentions commonly divide these into the following two kinds: intentions with respect to the future and intentions in acting. Personal intentions in acting are an important aspect of much human behavior. Indeed, a large class of words that are applied to such behavior in everyday life depend for their applicability on the presence of a certain intention. Thus, suppose someone's arm rises; this rising of the person's arm will only properly be described as his or her raising it, if he or she intends to raise it.

Although not all personal intentions with respect to the future would naturally be referred to as plans, one with a plan thereby has a personal intention with respect to the future. Such intentions bring an important element of organization into the personal lives of individual human beings.

Joint intentions can also be divided into intentions with regard to the future and intentions in acting. This discussion focuses on the former kind of joint intention.

Questions About Joint Intention

On what is probably the most natural reading of sentences of the form “We intend . . .” in English, they are not elliptical for sentences of the form “We both intend . . .” or “We all intend . . .” Rather, they ascribe an intention to us—a joint intention.

The question arises: What is a joint intention? Is there, indeed, such a thing as a joint intention? Is it really the case that two or more people, as opposed to each of a number of people, can have an intention? Some may think that this cannot be, as it suggests that, in addition to the minds of individual human beings, there are minds of another kind—group minds, if you like. And that may seem impossible. Since the late 1980s, there has been a fair amount of discussion of these questions by philosophers working in the Anglo-American analytic tradition.

Generally, these philosophers do not doubt that an intention can truly be ascribed to us as opposed to me on the one hand and you on the other. There is disagreement, however, on what our having an intention amounts to.

Some of the main issues that have emerged include the following: What is the relationship, if any, of a joint intention to the personal intentions of the participants? More broadly, what is the relationship of a joint intention to specific psychological states of the participants? Can joint intention be understood in terms of the attitudes of the participants at a given time without reference to their past history? Does there need to have been some form of communication between the parties? These and related questions concern the nature or constitution of shared intention.

Other questions relate to reasoning from a joint intention. If the participants in a joint intention wish to act appropriately, in the absence of other pertinent considerations, must their actions respect the shared intention? For instance, in the absence of other pertinent considerations, if we intend to fly to France from London tomorrow morning, is it incumbent on me not to buy tickets for a flight to France that leaves London in the evening?

Given that this is the case, am I also obligated to the others to respect the joint intention? That is,
does each participant owe the others such respect? What if the intention is to do something bad? Alternatively, what if one party comes to participate in the shared intention only because he or she is coerced into doing so by the other party?

**Schools of Thought on Joint Intention**

As to the relationship of a joint intention to specific psychological states of the participants, there are several schools of thought. This entry distinguishes between what might be called “correlated personal intentions,” “subjective we-intentions,” and “joint commitment” accounts, and makes note of the advantages and disadvantages of each.

**Personal Intentions Accounts**

A popular perspective on the matter centrally invokes appropriately correlated personal intentions of the participants. Such intentions are expressible with “I intend” and other similar expressions. One advantage of personal intentions accounts is that the idea of a personal intention is already familiar from action theory.

The personal intentions invoked in correlated personal intentions accounts have a variety of contents. Thus, Michael Bratman invokes personal intentions that we do such-and-such. More fully, Bratman’s earliest account of a joint intention to J posits a set of “interlocking” personal intentions and runs roughly as follows: I intend that we J; you intend that we J; I intend that we J by virtue of both my intention that we J and your intention that we J; and you intend likewise. Finally, it is common knowledge between the participants that they have these personal intentions. That means, roughly, that each knows of these intentions of each, and each knows this. In later work, Bratman added various further clauses.

One might also propose that a joint intention involves not personal intentions that we J, but rather personal intentions on the part of each to act as best he or she can to achieve their J-ing, given the actions of the other party. This is roughly in the spirit of work on the topic by Raimo Tuomela and Kaarlo Miller, among others.

**Subjective We-Intentions Accounts**

Rather than invoking a personal intention expressible with the words “I intend,” some philosophers, notably Wilfrid Sellars and, later, John Searle, have proposed that, on the contrary, our intending to do something is a matter of each of us being in a special psychological state, expressible by the words “We intend . . . .” The author may give the psychological state a special label, such as “we-intending.”

One problem with this proposal is that it is relatively obscure. It says little more than that those to whom a joint intention can correctly be ascribed are in a special psychological state, a state appropriate specifically to those who participate in a joint intention. Presumably this state involves some understanding of who “we” are, whether by description or by enumeration. Other than that, it is not clear what it amounts to. At the end of the day, the move from the question “What is a joint intention?” to “What is a we-intention?” does not take us very far.

The negative point made by those who advocate an account in terms of we-intentions as opposed to personal intentions, however, appears to be sound. There are significant problems with personal intentions accounts of joint intentions.

**Problems With Personal Intentions Accounts**

One problem with personal intentions accounts is that a joint intention to paint the house together soon, for instance, does not appear necessarily to involve personal intentions of the type to which personal intentions theorists appeal. One reason for saying this is as follows.

It seems that those who have agreed to paint a house together can immediately and truly say, “We intend to paint the house together,” simply by virtue of this agreement. And it is at least not obvious that those who have agreed to paint a house together tomorrow must each personally intend that they paint the house together tomorrow, or personally intend to do his or her part in their painting the house together, or something of that sort. Indeed, it is not obvious that each must have any particular personal intentions with respect to what “we” intend. One or more of them may have such intentions, of course, but the fact that they made the agreement in question does not seem to entail that they have such intentions. If no such personal intentions must be present when there is a joint intention, then clearly no personal intentions account will work.

Another problem with a personal intentions account is that those who participate in a given joint
intention tend to think and act as if, by virtue of one's participation in a joint intention, one owes the other participants actions that respect the shared intention. This is indicated by the kinds of rebukes and demands on one another in which participants engage. So one might say to another, “Why did you buy tickets for an evening flight? Our plan was to go in the morning.” Accounts of joint intention in terms of personal intentions have difficulty explaining such reactions, which appear to be based on the very existence of the joint intention or plan.

A further problem is the fact that participants in a joint intention understand themselves and the other parties not to be free to unilaterally alter or cancel the joint intention, absent special background understandings. Thus, a participant might appropriately respond to a rebuke such as the one mentioned in the previous paragraph with “Oh, I forgot” but not “Oh, I changed our plan!” To the latter response, the first party might object “But you can’t change our plan, not just like that!” Because one can alter one’s personal intention, just like that, this aspect of joint intentions may be impossible for a personal intentions account to capture.

Joint Commitment Accounts

An account of joint intention that provides an alternative to both personal intentions and subjective we-intentions accounts has been proposed by Margaret Gilbert. It appeals to more than the subjective states of the participants, as the we-intentions accounts do, and does not require personal intentions in favor of the joint intention, as the personal intentions accounts do. Gilbert has argued that when people have a joint intention, they owe each other conforming actions and are open to rebukes for nonconformity and demands for conformity.

Using the example of painting the house together, Gilbert’s account runs roughly as follows: Two or more people have a joint intention to paint the house together if and only if they are jointly committed to intend as a body that the house be painted by virtue of the coordinated activity of the two of them. It is understood that, in order that such a joint commitment be established, each has communicated to the other his or her readiness jointly to commit them all to emulate as far as possible a single creature that intends that the house be so painted. “Communication” is understood here in a broad sense. It can take place without face-to-face interaction.

In a special kind of case, the parties are jointly committed to accept decisions of a given person or body as to what they are jointly committed to intend as a body. The initial joint commitment establishes the person or body in question as having the authority to establish joint intentions for the parties.

What of joint intentions to do something bad and joint intentions such that one party or more of the parties have been coerced into participating in the joint intention? Gilbert argues that given such factors, the parties may well be able to argue that it is not appropriate to act in accordance with the shared intention, all things considered. Nonetheless, they can still be said to owe each other such action, in a particular sense of “owe.”

People owe one another actions in accordance with any joint commitment insofar as by jointly committing one another they have together imposed a constraint on each other, with respect to what they may appropriately do. To that extent they may together be said to have “put their dibs” on the action of each. So each may rebuke any other with respect to action that is not appropriate to the shared intention, in the name of them all.

Gilbert labels any set of jointly committed persons a “plural subject.” In using this label, she does not mean to imply that there is a group consciousness, or subjective state, distinct from the consciousness of each individual person. Hence, her account respects an important constraint that other theorists of joint intention have insisted on.

Her account also allows that those with a joint intention will appropriately form, where necessary, personal intentions that will support the joint intention, as when someone forms the personal intention to drive to the store for paint, in light of a joint intention to paint the house with another person, whom, he or she knows, will be buying other necessary items. Yet such personal intentions need not be present in every case where there is a joint commitment of the kind in question here. This respects the point made earlier as to the apparent nonnecessity of such personal intentions to the existence of a joint intention.

Margaret Gilbert

See also Action and Bodily Movement; Collective Action; Philosophy of Action
Further Readings


Know-How, Philosophical Perspectives

Any philosophical entry on know-how must begin with Gilbert Ryle, who was responsible for the modern distinction between knowledge-how and knowledge-that. Having set out the distinction, this entry will examine the use to which Ryle puts it in his critique of René Descartes, before considering the role it plays in contemporary philosophy, particularly in the debate over physicalism.

Knowledge-How Versus Knowledge-That—Ryle’s Distinction

To a first approximation, the distinction we inherit from Ryle is between what might be termed intellectual knowledge-that, which is propositional in nature, and practical knowledge-how, which may be understood in terms of having abilities to do certain things. Consider a project to learn all about bicycles. There is a lot of knowledge-that to be gained about bicycles: One can learn that bicycles have two wheels, that they balance along their longest axis, that one must pedal while remaining balanced to keep the bicycle upright, and so on. (The phrases after each that in the previous sentence express facts or propositions—hence, knowledge-that is propositional knowledge, knowledge of propositions.) But still, it is fairly obvious to most that even someone who has learned all the propositions there are to know about bicycles will not thereby be gifted with the ability to ride a bicycle. We might expect such

a cycling know-it-all to fall flat on his face when attempting his first unaided ride. What the know-it-all lacks, the story goes, is know-how: He does not know how to ride a bicycle. Alternatively, we might say that having learned academically all about bicycles, this learner has yet to acquire the practical ability to ride bicycles. Further, whereas time in the classroom arguably suffices for acquiring all the knowledge-that about bicycles that our student has come to possess, we know that acquiring, in addition, the knowledge how to ride a bicycle will require some first-hand experience—some practice in bike riding on the student’s part.

Ryle’s Critique of Descartes

Ryle introduced the previously described distinction in his important book The Concept of Mind, as part of his campaign against the then (Ryle is writing at the start of the 1950s) orthodoxy of a Cartesian conception of mind and mentality. René Descartes had proposed his dualism: the idea that the mind is a nonphysical object separate from, though connected to, the physical body and, consequently, not observable or measurable using conventional scientific methods. Ryle strongly objected to the notion that mental processes go on in secret: hidden and unobservable as a matter of principle. He posed the following dilemma for what he labeled Descartes’ “intellectualist legend,” the doctrine that all intelligent bodily action (as opposed to mere reflex) is preceded by rational thought—the consideration of a relevant proposition—that effectively plans the next move for the body. Ryle observes that planning a
bodily movement, considering propositions, is itself an operation, an action of sorts. Now if the act of proposition-consideration itself were unintelligent, it would be hard to see how it could confer intelligence on the bodily movement that it preceded. So clearly, the act of considering the relevant proposition prior to originating an intelligent bodily action had itself better be intelligent. But what is it to be an intelligent act according to Ryle’s Cartesian intellectualist legend? It is to be preceded by a mental operation of considering a relevant proposition. In which case the mental planning procedure that originated the bodily action in this case itself needs to be preceded by a mental act of planning, an act which, in turn, must be intelligent and thus preceded by a further contemplative act, and so on. The dilemma that arises for the Cartesian, then, is this: Either an intelligent action can never get started, for it must always be preceded by a further mental act of planning, or some actions are intelligent without being preceded by mental operations, operations Ryle considers appallingly occult on Descartes’ conception of mind and its processes. This reasoning led Ryle to posit a distinct, nonintellectual, practical genre of knowledge, that is, knowledge-how, knowledge, in some sense, of the body.

Know-How Versus Know-That in Modern Philosophy

The know-how/know-that distinction as proposed by Ryle is of philosophical interest in its own right, as well as for the part it plays in the Rylean critique of Descartes. Nowadays the distinction is more familiar to philosophers because of its employment in another controversy concerning the philosophy of mind. Frank Jackson tries to disprove physicalism—the doctrine that all that exists is physical matter and its combinations—with an argument that centers on an omniscient color scientist, Mary. Mary knows all the scientific facts relating to human color vision, despite never having seen color. Living her life hitherto in a monochrome laboratory, Mary has extensively studied red-seeing subjects in the outside world, thus accumulating her stock of scientific knowledge. One day, Mary is shown a red rose. Jackson invites us to agree that Mary learns something important in this encounter: the fact of what an experience of redness is like, qualitatively, for the person undergoing it. In which case Mary’s prior stock of physical-factual knowledge concerning color vision was incomplete, and it follows, claims Jackson, that physicalism is false. There are some nonphysical features of the world to be learned about.

One of the first and most enduring replies to this provocative argument draws on Ryle’s distinction. David Lewis, following Lawrence Nemirow, asserts that what Mary learns when she sees red is not any new fact, but a set of new abilities—in other words, she gains no knowledge-that, only knowledge-how. Specifically, Lewis proposes that Mary acquires the abilities to recognize, remember, and imagine the experience of redness through her meeting with the rose. However, because Mary learns no new fact, as there is no new proposition that Mary now knows, Lewis claims, physicalism is not threatened by Mary’s increase in knowledge. She learns how to do some new things, but there is no new—and so nonphysical—fact that she discovers.

Lewis’s argument has brought renewed attention on Ryle’s distinction, mostly in the service of discussions concerning mind, but also independently. Some remain dubious about the distinction. An article by Jason Stanley and Timothy Williamson makes a good case for the assimilation of knowledge-how to knowledge-that. But Stanley and Williamson’s thesis, that knowing how to ride a bicycle (for example) is really knowing, of some manner of riding a bicycle, that this is a way to ride a bicycle, is itself widely rejected. Thus, the controversy rumbles on.

Sam Coleman is also sympathetic to the view that know-how, or ability knowledge, might reduce to knowledge of fact. He notes that abilities, such as the ability to ride a bicycle, depend on knowing what certain sensations feel like—in the case of cycling, the would-be rider must, crucially, become well acquainted with the sensation of balancing on her bicycle in order to gain the ability. But if knowing what sensations feel like is factual knowledge—as Jackson suggests—then know-howability knowledge will be partly constituted by knowledge-that. If this is so, then Ryle’s distinction may collapse, and with it Lewis’s popular objection to Jackson’s “knowledge argument.”

Sam Coleman

See also Knowledge by Acquaintance; Mind-Body Problem; Physicalism; Reductive Physicalism
Further Readings


Knowledge Acquisition in Development

Childhood is a period of remarkable knowledge acquisition—unparalleled in human learning. In the first five years of life, children are transformed from helpless infants with virtually no understanding of the world around them to articulate students with a rich understanding of time, space, and number; an ability to organize objects and animals into category hierarchies; a capacity to infer cause and effect; and sensitivity to the emotional states of others. In the next dozen years, children are further transformed, as they engage in complex reasoning in domains including politics, art, history, moral judgments, and science. As impressive as these intellectual achievements are, they are tempered by the observation that children's knowledge acquisition is also shaped and constrained by pervasive reasoning biases. Children find certain kinds of concepts difficult to acquire and easy to misunderstand (such as evolutionary theory, fractions, or irony), they hold firm misconceptions in nearly every domain studied, and they insufficiently consider the available evidence when making decisions. These two themes (remarkable learning and persistent biases) may at first seem contradictory, but they are not. Remarkable learning often occurs because of early biases.

The study of knowledge acquisition sheds light on several classic theoretical debates in psychology, including the following: What is the interplay between innate capacities and environmental experiences? What is the role of domain-general processes and processes specific to particular domains of knowledge? How much continuity or discontinuity is found across development? What are the forces (evolutionary, social, neurological) that contribute to developing conceptual systems? Although each of these topics remains the focus of active debate, research over the past 50 years provides rich insights into these core questions.

This entry discusses four questions that have been a focus in studying knowledge acquisition in development. Traditionally, scholars have focused on three primary aspects: process (what are the mechanisms that enable learning to take place), content (what do children know), and structure (how is knowledge organized). More recently, researchers have emphasized a fourth aspect: the role of context and culture (how is learning influenced by the social and cultural context in which it takes place). Each of these approaches has theoretical, practical, and educational significance. For example, understanding the process of how children encode, retain, and retrieve memories has implications for improving learning skills as well as establishing procedures for child witness interviews. Determining the content of children's knowledge about illness can help when counseling a child with a seriously ill sibling. Discovering the structure of children's knowledge of causality has theoretical significance for understanding whether scientific reasoning is continuous or undergoes considerable restructuring over development. Finally, examining the contextual effects of parent-child conversations on a child's language and conceptual development has important implications for administering advice and interventions to parents.

Knowledge Acquisition as an Active, Constructive Process

Theories of knowledge acquisition have historically distinguished empiricist from nativist views. Empiricists suggest that children enter the world as “blank slates,” without any innate mental structures
beyond basic sensory capacities and domain-general learning mechanisms. Thus, knowledge is viewed as built up primarily from experience. In contrast, nativists suggest that infants are endowed with innate knowledge and abilities that unfold over time. In contrast to both of these views, a constructivist approach (epitomized by Jean Piaget) proposes that children actively construct their knowledge. On this view, children enter the world with basic capacities to organize information in the world into mental structures and then elaborate on these structures as new information is encountered. In this way, children are like scientists, forming theory-like representations of the world and interpreting new information in light of existing theories.

Piaget suggested that knowledge development involves qualitative change, with children proceeding through four developmental stages marked by distinct modes of thinking: sensorimotor (birth to roughly age 2 years), pre-operational (roughly 2–6 years of age), concrete operational (roughly 7–12 years of age), and formal operational (roughly 12 years on). However, it is now widely acknowledged that a strict stage view of knowledge development is incorrect, as it underestimates children’s early capacities and overestimates the rationality of adolescents and adults. Early knowledge is much richer than previously thought. Young children understand certain concepts before they can demonstrate them (also known as a competence-performance gap). For example, infants younger than 9 months of age fail to search for an object that is covered with a cloth, seeming to indicate an implicit belief that “out of sight is out of mind.” However, careful experiments that track infants’ gaze and reaching behaviors indicate awareness of hidden objects as young as 3 to 4 months of age. Contemporary researchers have developed numerous other implicit and subtle measures (e.g., rate at which infants suck a pacifier as a measure of interest in a stimulus, sequential order of manual object exploration as a measure of categorization, neuroimaging techniques as measures of the role of attention and control in cognitive tasks), which reveal that basic cognitive capacities are in place early in development.

There are several contemporary views of cognitive development that differ from Piaget’s view but build on his insights. For example, information-processing theories emphasize that development involves domain-general changes, such as increased processing speed or working memory capacity. On this view, developing brain capacities and increasing knowledge of the world (expertise) contribute to general developments in capacity, reasoning strategies, and performance. Structural theories posit that development involves acquiring the ability to reflect on early representations, with knowledge proceeding from largely procedural and unconscious to more explicit and deliberate. Theory theories suggest that children are born with innate, albeit rudimentary, knowledge about the world that constrains how they process and interpret information they encounter, but that important reorganizations of knowledge take place with development. For example, preschoolers distinguish between material and immaterial entities (e.g., objects vs. thoughts) and understand some principles about material substance (e.g., that matter occupies physical space), but, between ages 4 and 12, their concept of “matter” undergoes considerable reorganization as they incorporate more complex principles (e.g., that all matter has weight, regardless of its size or density).

Content and Structure of Early Knowledge

Children’s naïve “theories” about the world differ from scientific theories but are similar in the following respects: They presume a domain-specific ontology (e.g., “objects” are units in naive physics; “animals” are units in naive biology); they are constructed from evidence; they generate predictions and causal explanations; they posit unobservable theoretical entities (e.g., gravity) to account for observable phenomena (e.g., objects falling); they are coherent; and they are defeasible in the face of counter-evidence. Three foundational domains in which children construct causal theories include physics, psychology, and biology. In each of these domains, there are impressive early capacities as well as considerable changes across development.

Fundamental elements of a naïve theory of physics are evident early in infancy. Infants represent objects as solid, bounded entities whose behavior accords with certain physical regularities. They expect objects to continue to exist when out of sight, and they expect them to move lawfully (e.g., not through obstructions). With age, children acquire richer understandings of physical and mechanical principles. For example, they are not born with an understanding that objects obey the laws of gravity, but begin to develop expectations about gravity by about 2 years
of age. Further changes in physical understanding take place even into adulthood, as adults struggle to overcome an intuitive physics in which they have pre-Newtonian beliefs about the physics of everyday objects (e.g., incorrectly predicting that a ball rolling out of a curved tube will continue on a curvilinear, rather than a linear, trajectory).

A naïve theory of psychology, also known as a “theory of mind,” also has precursors in infancy. By 5 months, infants interpret behaviors of animate entities as goal directed and distinguish between intentional and accidental actions. By preschool age, children distinguish between mental and physical entities (e.g., thoughts vs. actual objects), link perception with knowledge, and see people as having beliefs and desires that are linked to their actions. Richer understandings of the links among beliefs, desires, and actions emerge across development. For example, not until age 6 do children appreciate that differences in preexisting expectations differentially influence how people interpret ambiguous events.

Knowledge of the biological world entails classifying living things and reasoning about biological processes, such as growth and reproduction. There is debate about whether children’s biological knowledge constitutes a theory; however, at the very least, by preschool age, children exhibit certain key understandings. For example, they distinguish between biological and inanimate entities and appreciate that biological processes only occur with the former. Less is known about biological knowledge in younger children; however, infants expect animate objects to exhibit self-initiated movement and inanimate objects to require external force, which is a likely precursor to a naïve biological theory.

As children construct knowledge systems, they also reveal systematic biases in how they interpret and incorporate new information. These include an essentialist bias (assuming that categories have an underlying reality), a teleological bias (assuming that all entities and events have a purpose), and causal determinism (assuming that all entities and events have a cause), among others. These biases are particularly evident in early childhood, but they may also persist into adulthood.

**Context and Culture**

Until recently, the focus in cognitive development was primarily on characterizing children’s knowledge, and relatively little attention was paid to the contextual nature of the input. For example, Piaget provided the example of a child discovering basic principles of mathematics by rearranging and re-counting a set of pebbles. However, learning is not entirely a solitary act; instead, it is embedded in social and cultural understandings. Much of children’s knowledge is derived not from their direct interactions with the environment but rather from the testimony of knowledgeable others. Studies of theory of mind tell us that learning often requires attending to others as a crucial source of information. Social transmission is also a mechanism for transmitting scientific concepts (germs, shape of the earth), natural categories (tomatoes are fruit), social concepts (ethnicity, personality traits), and supernatural concepts (God, witchcraft). The typically developing child interprets and evaluates the surrounding social input, and disruptions to these capacities can be devastating (as with autism). Cultural factors also play a key role. In a long tradition influenced by Lev Vygotsky, cultural psychologists have concluded that cultural contexts significantly influence children’s learning. Finally, comparative studies with humans and nonhuman species suggest that certain forms of social learning—imitative learning, instructed learning, and collaborative learning—may be unique to humans (or if not unique, then at least particularly well developed). Humans are the preeminent species that create culture and cultural artifacts. Thus, human knowledge development is uniquely influenced by social and cultural factors.

**Conclusions**

The study of knowledge development in childhood provides a particularly fruitful approach to understanding human intelligence. Much of what makes our species distinctive is our capacity to acquire information (from observing the world and learning from others), to organize that information into explanatory and predictive theories, and to reorganize knowledge in the face of new and unexpected evidence. These capacities are rooted in fundamental processes that are evident early in infancy and continue to grow and flourish with maturity and experience.

Elizabeth A. Ware and Susan A. Gelman

See also Concepts, Development of; Folk Psychology; Language Development; Representations, Development of; Social Cognition
Knowledge by Acquaintance

Knowledge by acquaintance is knowledge of an object that depends solely on one’s acquaintance with the object. By contrast, knowledge by description is knowledge of an object that depends on one’s knowledge of descriptive truths about the object. The distinction between knowledge by acquaintance and knowledge by description played an important role in the philosophy of Bertrand Russell, and it continues to inform much contemporary work in epistemology and the philosophy of language and mind. This entry is divided into four sections: (a) the acquaintance relation, (b) the objects of acquaintance, (c) the role of acquaintance, and (d) contemporary work on acquaintance.

The Acquaintance Relation

Acquaintance is a relation that holds between subjects and the objects of their acquaintance. But what kind of relation is it? It is sometimes said that the nature of acquaintance cannot be known by description, but only by acquaintance—in other words, one must be acquainted with acquaintance in order to know what it is.

Acquaintance with an object is usually defined as a conscious state of direct and unmediated awareness of an object. This follows Russell’s (1912) definition: “We shall say that we have acquaintance with anything of which we are directly aware, without the intermediary of any process of inference or any knowledge of truth” (p. 25).

According to Russell, there are various different forms of acquaintance, including not only perception, but also memory, introspection, and conceptual awareness of universals. Nevertheless, perception is the least controversial example: In perception, we are acquainted with objects. But what are the objects of our perceptual acquaintance?

The Objects of Acquaintance

According to common sense, we are acquainted in perception with ordinary physical objects, including tables, chairs, and other people. According to Russell, however, when I see a table, I am not directly acquainted with the table, but rather with mental objects, which he calls sense data. Therefore, I do not know the table by acquaintance, but merely by description, as “the physical object which causes such-and-such sense data.” On this view, expressions used to refer to physical objects, including demonstratives and proper names, are really descriptions in disguise.

Russell uses a version of the argument from illusion to support his view. He argues that, when I view a tilted coin and seem to see an elliptical object, there is in fact an elliptical object that I am seeing. However, this object is not the coin, which is circular, but rather a mental sense datum. Critics of the argument deny that, in cases of illusion, if it seems that I am seeing an object that has certain properties, there is in fact an object that I am seeing, which has the relevant properties. Thus, there is no elliptical object that I am seeing; rather, I am seeing the circular coin, but its apparently elliptical shape is illusory. However, the problem of giving an adequate account of illusion remains one of the central problems in contemporary philosophy of perception.

The Role of Acquaintance

Acquaintance is central to Russell’s theory of knowledge and his theory of conceptual thought. He draws a distinction between knowledge of things...
Knowledge by Acquaintance

and knowledge of truths, which is best understood in terms of the distinction between our conceptual ability to think about things and our epistemic ability to know truths about those things. According to Russell (1912), acquaintance plays a foundational role in explaining each of these abilities: “All our knowledge, both knowledge of things and knowledge of truths, rests on acquaintance as its foundation” (p. 48).

First, acquaintance plays a foundational role in Russell’s theory of knowledge. He argues that all of our knowledge of truths depends ultimately on acquaintance, which is foundational in the sense that it does not itself depend on any knowledge of truths. Acquaintance with an object enables us to know descriptive truths about the object, but this descriptive knowledge depends on acquaintance, rather than vice versa.

Second, acquaintance plays a foundational role in Russell’s theory of conceptual thought. He argues that we can think about an object only if we know which object we are thinking about. Either we know the object by acquaintance or we know it by description, in which case we must be acquainted with the properties in terms of which it is described. This is the rationale for Russell’s (1912) principle of acquaintance, which states: “Every proposition which we can understand must be composed wholly of constituents with which we are acquainted” (p. 58).

Contemporary Work on Acquaintance

Few contemporary philosophers endorse all aspects of Russell’s philosophy, but many follow his example in finding an important role for acquaintance in the theory of knowledge and conceptual thought. A common strategy is to identify conscious states that play the role of Russell’s notion of acquaintance in grounding our conceptual and epistemic abilities. For example, John Campbell argues that conscious visual attention to an object enables one to grasp demonstrative concepts of the object and to know truths about its visible properties. Similarly, David Chalmers argues that introspective attention enables one to grasp phenomenal concepts and to know truths about the phenomenal properties of one’s phenomenally conscious mental states.

Since Russell, there has been widespread agreement on the importance of a distinction between thinking of an object by description and thinking of an object in a more direct way that exploits a descriptively unmediated relation to the object. Indeed, the term acquaintance is sometimes used as a mere placeholder for whatever relation it is that enables one to think nondescriptive thoughts about an object. From this perspective, however, it is a substantive question whether Russell was entitled to assume that the role of acquaintance could be played only by states of conscious awareness.

For example, Gareth Evans introduces a category of information-based thought, which exploits information that is causally derived from an object. According to Evans, the role of Russell’s notion of acquaintance is played by the information link with an object that is provided in perception, memory, and testimony. However, it is a focus of contemporary debate whether the information provided by perception, memory, and testimony must be conscious to play the role of Russell’s notion of acquaintance.

In sum, Russell’s work on acquaintance raises important questions about the role of consciousness in grounding our conceptual and epistemic abilities. These questions are central to much current work in epistemology and the philosophy of language and mind.

Declan Smithies

See also Descriptive Thought; Know-How, Philosophical Perspectives; Object-Dependent Thought

Further Readings


Language Development

Language development is the process by which children come to understand and produce language. This entry focuses on the period between birth and 5 years of age. The entry presents theoretical perspectives on language development, followed by a review of the biological basis of language. The language acquisition process is then described for various components of language, with a focus on typically developing children learning one first language; brief sections also discuss bilingualism and atypical language development.

Historical and Theoretical Perspectives on Language Development

In the first half of the 20th century, the field of language development was dominated by studies documenting the normative course of language acquisition. In the late 1950s and 1960s, the linguist Noam Chomsky published several works that took the field in different directions. The first was a review of B. F. Skinner’s book Verbal Behavior. Skinner applied behaviorist theory to language development, arguing that it results from processes such as imitation and reinforcement. Chomsky’s review was a rebuttal of the application of behaviorism to language and suggested that language development is much more complex. Chomsky followed with his theory of innate linguistic knowledge in which he claimed that there is a language acquisition device in the human brain that contains innate knowledge of the structure of language, also known as universal grammar.

Chomsky’s theory is a nativist view of language development as it claims that children have preexisting knowledge of language, whereas Skinner’s behaviorist approach is an empiricist view that all knowledge of language comes from experience. The extreme empiricist view is not popular today. The interactionist perspective provides a more moderate approach to language learning by highlighting the importance of experience in language development while acknowledging the existence of brain structures that support language development. The main difference in the nativist versus interactionist perspectives lies in the importance they place on experience. Further, in their emphasis on experience, interactionists tend to see the child as playing more of an active role in language development than do nativists, who see the development of language more as something that happens to the child in a predetermined way. Although much has been learned over the past 50 years about child language development, there are still many unanswered questions.

Biological Basis of Language

Almost all humans learn to talk. From birth, human beings are biologically prepared for language. A specialized vocal tract helps humans produce language, and the position of the larynx and properties of the lips and tongue make rapid sounds easy to produce. Upright teeth, while not necessary for eating, allow for the production of certain sounds such as /s/ and /θ/. These characteristics are unique to humans. The
vocal tract is not the only human characteristic that makes language production possible; specific parts of the brain work in tandem with the vocal tract to produce language. For 85% of the population, the left hemisphere is dominant for language processing. This is known as functional asymmetry, in which one hemisphere in the brain plays a different or larger role than the other for a specified function.

The classical model of lateralization argued that specific regions of the left hemisphere are particularly important for language functions. Based on symptoms of Broca’s aphasia, Broca’s area was originally thought to be responsible for language production; however, the underlying disorder is now thought to involve grammar and phonology in both comprehension and production. Similarly, lesions to Wernicke's area, a left-hemisphere area originally thought to be involved in written and spoken language comprehension, are now thought to underlie semantic and conceptual deficits in both production and comprehension. While the right hemisphere can comprehend some language, it is especially involved in aspects of pragmatics, prosody, and discourse comprehension. Both hemispheres are involved in semantic processing, but in different ways. For instance, there is evidence that the left hemisphere activates the main meaning of a word in a particular context, while the right hemisphere activates a broader range of meanings. The role of each hemisphere in language processing, and how the hemispheres interact, is still being examined.

The human brain also appears to be quite plastic in terms of language acquisition and development. If left-hemispheric injury occurs early in life, the right hemisphere sometimes compensates for the injured area’s functions; the ability of the right hemisphere to take over may decline with puberty. Researchers take this finding to argue for a critical period, or a time (often before age 8–13) during which language is more readily learned.

**Development of Pragmatic, Phonological, Lexical, and Grammatical Skills**

Language can be broken down into several components that develop concurrently in the child: pragmatics, phonology, lexicon, and grammar. The following sections provide the typical developmental progression within each component across early childhood.

**Pragmatics**

Pragmatic development refers to the understanding of how to use language appropriately and to serve different communicative functions and intents (e.g., to direct someone’s attention or ask a question). Children can understand and express communicative intents before they can use spoken language productively. One way they do this is through communicative gestures such as pointing. For example, a 1-year-old who wants an object that is out of reach might request the object by pointing to it and then looking at an adult. Alternatively, if an adult is taking an object away from an infant, the child might protest by getting upset and shaking the head to mean “no.” These early communicative behaviors suggest that the child has developed intentionality, which, in this context, is the ability to communicate intent (e.g., in the previous cases, to make a request and to protest). Intentionality is found to develop in preverbal infants around 10 months of age. Detailed studies of children’s early pragmatic development suggest that 1-year-olds have a repertoire of between 5 and 30 communicative intents, whereas 2-year-olds have a repertoire of between 50 and 90 intents. Thus, as children increase in their formal abilities to use language productively, they also increase in the communicative purposes for which they can use language.

Pragmatic development continues across childhood and encompasses the development of conversational skill. Although young children can use language for many communicative acts by age 2, the ability to engage in conversation requires skills that often do not develop until years later. For example, children need to understand the conversational rules of turn taking and topic relevance. Understanding culturally relevant rules of politeness also falls under the realm of pragmatics.

**Phonology**

Phonological development involves understanding how to distinguish between, and produce, the sounds of the adult language. Phonemes are the meaningfully different sounds in a given language, such as /g/ and /d/ in English. From birth, infants can distinguish between phonemes when they are presented in isolation using habituation/dishabituation procedures such as the high amplitude sucking paradigm. In this paradigm, infants will slow down
their pacifier sucking once they are bored with a stimulus (e.g., the sound “pa”). If the infant’s sucking then increases on presentation of a new phoneme (e.g., “da”), it is an indication that the infant detects the difference in sounds. Interestingly, after about 9 months of age, infants have a harder time distinguishing between phonemes in languages other than their own. Thus, language experience influences language perception, as it is easier for infants to perceive differences in phonemes they hear regularly.

The first sounds infants produce are vegetative sounds such as crying and burping. Around 6 to 8 weeks, infants begin cooing or producing long drawn-out vowel-like sounds of contentment. Between 4 and 7 months, babies engage in vocal play and increase their repertoire of sounds. Vowels are produced first because they are easiest to produce physiologically, followed by consonants formed in the back of the mouth (/g/), and then by consonants formed in the front of the mouth (/m/). Between 6 and 9 months, children start babbling, first by producing reduplication of true syllables from their language, also known as canonical babbling (e.g., /nana/), and later by working in a larger range of phonemes. During this period, children also produce jargon where they string together sounds with the melody and intonation of their language. Indeed, studies have found that parents exaggerate their intonation when talking to infants as compared with talking to adults. As is found with perception of sounds, after 6 months of age, the sounds infants produce start to resemble the sounds of the language(s) they are exposed to rather than other languages.

Children’s first words often have a simple syllable structure. During the second year of life, when children are limited in their repertoire of phonemes, they find creative ways to transform sounds to make them easier to produce. For example, they use what is called reduplication of syllables to transform bottle to /baba/ or delete syllables in words, such as transforming banana to /nana/. They often mispronounce words during this period as well. By age 3 years, speech is more intelligible, and by age 7, children sound adultlike in their phonology. Children show some signs of phonological awareness—that is, awareness of their phonological abilities—as early as age 2, when they play with sounds and appreciate rhyme. As children get older, this awareness grows and they can correctly answer questions such as “say cat without the /c/ sound.” Phonological awareness is an important precursor for learning to read.

**Lexicon**

Words are difficult to learn because they are arbitrary symbols with no inherent relation to their referents. The ability to map a word to its referent (e.g., to know that when they hear “dog,” it is referring to the four-legged animal) is crucial to lexical development. How children do this is a topic of much debate. Nativists would argue that children are born with knowledge of how words work. Other psychologists explain the mapping problem using a variety of constraints. For example, the whole object constraint suggests that children first assume words to refer to a whole object, rather than its parts. Social-cognitive psychologists assert that children solve the mapping problem by focusing on the intentions of others and the meanings they are trying to convey. Finally, researchers with more of a domain general perspective would argue that other cognitive processes, such as attention and memory, can help explain word learning. Thus, the process of word learning is less clear-cut than the facts of word learning at different ages.

As with other aspects of language, comprehension precedes production in lexical development. Children tend to recognize their own name by 5 months, and by 10 months they can comprehend between 10 and 150 words as well as simple phrases, yet they do not often produce their first words until they are 10 to 15 months old. Children reach a productive vocabulary of 50 words between 15 and 20 months. In children learning English as well as many other languages, on average approximately half of the first 50 words produced are nouns (e.g., ball, mama), yet there is variation across children. However, in some languages such as Mandarin, there is less of a noun bias in children’s early lexicons. Differences across languages may be due to the structure of the languages themselves in how nouns and verbs are used, and to cultural differences in parenting practices such as labeling objects.

By age 2, children can produce between 50 and 550 words. Some of these large individual differences in vocabulary development can be explained by phonological memory as well as environmental factors. For example, studies consistently find a positive relation between the variety of vocabulary words parents
use with children and children's vocabulary size. As toddlers, children often produce unconventional word-meaning mappings. Overextensions refer to using a word more broadly than is appropriate such as using “dog” to refer to all four-legged animals, whereas underextensions refer to using a word more narrowly, such as using “dog” to refer only to German shepherds but not to collies.

As children's vocabulary size increases during early childhood, so does the depth of their vocabulary in that they gain greater understanding about specific words and their meanings and appropriate uses. Vocabulary acquisition is by no means complete at age 5. In fact, school-age children have been shown to increase their vocabularies by approximately 3,000 words per year.

**Grammar**

Grammatical development includes an understanding of morphology and syntax. Morphology consists of the rules for combining morphemes to create words. Morphemes are the smallest unit of meaning in language. Free morphemes are words that stand alone (e.g., *dog*) whereas bound morphemes cannot stand alone (e.g., plural *s*). Morphemes can be combined by (a) compounding, or combining two or more free morphemes (e.g., *houseboat*); (b) derivation, or combining a bound and free morpheme to change the meaning (e.g., *teacher*); or (c) inflection, or combining a bound and free morpheme without changing the meaning or grammatical category of the word (e.g., *dogs*).

Evidence that children acquire morphological rules comes from observing children's productive errors (e.g., *falled*) as well as from experimental studies. A famous early study in 1958 by Jean Berko used nonsense words to test children's morphological development. An experimenter first presented children with a picture and labeled it with a nonsense word such as *wug*. The experimenter then presented children with another picture and said, “Now there is another one, there are two of them, there are two ____?” If children could correctly answer “wugs,” this was evidence that they had learned the morphological rules because they had never heard the word *wugs* before in their language. The preschoolers in her study (the youngest subjects) performed well on these tasks. Another measure of children's early productive morphological development is a count of the mean length of utterance (MLU) measured in morphemes. Roger Brown developed stages of grammatical development based on the MLU of children he was observing longitudinally. He transcribed all their utterances, counted how many morphemes were in each utterance, and then averaged the total utterance length per child at each age. On average, children between 16 and 31 months are shown to have MLUs between 1.0 and 2.0. MLUs increase to between 2.0 and 3.0 for children between 21 and 41 months, and MLUs of 4.0 or more are reached by children around 3 years old or older.

Syntax is the component of grammar that governs the ordering of words in sentences. Studies show that as early as 16 to 18 months, children can comprehend meaning carried in word order. As children begin to produce language, they go through phases of syntactic development, which becomes increasingly complex with age. In the two-word phase of productive grammatical development, children tend to use the same words in different combinations to express possessives such as “my chair” and “my book” and descriptives such as “pretty doll” and “pretty dress.” Three-word speech combinations tend to be telegraphic at first in that they omit morphemes (e.g., “put it table”); at around 2 to 3 years, children fill in these obligatory morphemes. At around 3.5 years, when children reach an MLU of 3.0 to 4.0, they also start to produce passives as well as complex sentences. However, some more complex aspects of syntax such as anaphora, or the understanding of how pronouns refer to their referents in a sentence, are not mastered until middle childhood.

**Bilingual Language Acquisition**

Recent estimates suggest that almost 50% of the world's population is bilingual, and by 2030, 40% of American school-aged children will be English-language learners. The study of bilingualism is important given these social circumstances, as well as to learn about how bilingualism develops and varies across children. Bilingualism develops either sequentially, when one language is introduced after the acquisition of the first language, or simultaneously, when both languages are learned in tandem.

Although monolinguals and bilinguals learn language in some similar ways, bilingualism has been shown to slightly alter the course of development of certain linguistic and cognitive processes. Children in both monolingual and bilingual environments begin canonical babbling at approximately the same time. Bilinguals, however, have smaller vocabularies in each language than their monolingual peers;
lexical knowledge appears to be distributed across the two languages. Bilinguals, however, are found to show greater metalinguistic awareness, phonological awareness, grammatical awareness, and cognitive control. Proficiency in second language learning depends on several factors. First, individuals who are exposed to a second language earlier in childhood are better able to speak without an accent and ultimately master certain grammatical structures, because they have more opportunities to speak. However, receptive vocabulary, translation, and story comprehension seem to be skills that are better honed when second language acquisition occurs at an older age.

Atypical Language Development

Studying children with atypical trajectories of linguistic development is important for understanding how various abilities contribute to language acquisition as well as to help develop new interventions for language disorders. Typical language development relies on exposure to linguistic models in the environment. Deaf children born to hearing parents are not exposed to these models and must use alternative ways to acquire language and communicate with others. Approximately 1 in 1,000 children are born with severe hearing loss, yet if they are exposed to sign language from birth, deaf children demonstrate a typical linguistic progression using sign. These children typically babble manually rather than vocally, and they produce their first sign before or at the same time as typically developing children produce their first words.

Studying language disorders can also help researchers better understand typical language development, specifically with regard to the extent to which language and other cognitive processes are interrelated. For example, children with Down syndrome show general cognitive defects and are also late to babble and begin talking; they produce toddler-like phonological patterns into adulthood. Thus, cases of Down syndrome provide some evidence that linguistic and cognitive processes operate in tandem, as both are delayed in these children. On the other hand, some children who have impaired cognitive development do not have impaired language development. This provides evidence that cognitive and linguistic functioning may be separate processes. Individuals with Williams syndrome, for example, have IQs that are similar to those with Down syndrome but have stronger language abilities. In contrast, children with another language disorder, specific language impairment (SLI), do not display cognitive deficits. SLI usually results in language production 1 year behind and comprehension 6 months behind typically developing children. SLI seems to be hereditary, as 20% of children with SLI have relatives with language difficulties.

Children with autism spectrum disorders (ASD) provide an example of the social nature of language development. Lower functioning children with ASD show delayed linguistic development, and some of these children never acquire language. Higher functioning children with ASD, however, display similar syntactic abilities to typically developing children but struggle with communicative competence. Toddlers with ASD show little interest in others, do not initiate joint attention, and rarely produce pointing gestures. Later, pragmatic development is also impaired in children across the autism spectrum.

In sum, in the first 5 years of life, children make great gains in oral language skills. These abilities are found to contribute to the acquisition of reading and writing during school. Further, language development is by no means complete when children enter school; rather, it is a lifelong process, as even adults continue to build their vocabularies and gain greater understanding of language over time.

Meredith L. Rowe and Kathryn A. Leech

See also Bilingual Language Processing; Bilingualism, Cognitive Benefits of; Dyslexia, Developmental; Gesture and Language Processing; Heritage Language and Second Language Learning; Innateness and Parameter Setting; Language Development, Overregulation in; Word Learning

Further Readings


The study of language development has revealed many interesting findings and has contributed to the field of language acquisition and, more generally, to linguistics and psychology. Of particular importance is the emergence of certain intriguing patterns that are characteristic of early language development. This entry will discuss one particular feature of language development, namely, overregulation. Overregulation in language development occurs when children apply linguistic generalizations in contexts where that would not be deemed appropriate for adult speakers.

### Linguistic Background: The Case of the Past Tense

For instance, in English, verbs can be classified into two main types: regular and irregular. Regular verbs are so-called as their past tense forms are created by the addition of the suffix -ed to the stem (e.g., walk—walked). On the other hand, the irregular verbs cannot be summarized under one “rule” as can be done for the regular verbs. As such, the relationship between the stem and the past tense form is more arbitrary (e.g., go—went).

### Properties of Overregulation

#### U-Shaped Curve

At some point during language development (around the age of 2 years), children begin acquiring past tense formations. Although they produce both regulars and irregulars more or less accurately in the early stage of past tense acquisition, as they begin to acquire more verbs, they start producing errors, especially with regard to irregular formation. It is at this point during language development when children incorrectly inflect irregular verbs in a regular manner. For example, they may produce *goed* instead of *went*, or *thinked* instead of *thought* (an asterisk indicates ungrammatical forms). This developmental pattern is commonly referred to as a U-shaped learning curve. The irregular past tense forms are said to undergo this U-shaped learning curve because, during early acquisition of the irregular past tense, children produce inflected irregulars correctly. They then go through a phase where they overgeneralize the regular formation in which -ed is simply added to the verb stem. This overregulation of the regular formation to irregular verbs results in irregulars being produced incorrectly. In the final phase of the learning curve, children’s performance on irregulars increases as they learn to associate the correct irregular past tense form with its stem.

### Explaining Overregulation

It is worth examining why this phenomenon occurs. Recall that the past tense can be inflected regularly or irregularly. In English, many irregular verbs are highly frequent (i.e., they occur quite often). This could explain why the irregular forms are produced accurately at first. However, although individual tokens of irregular verbs are high in frequency, there are many more regular verbs than irregular verbs in English. This difference in type frequency (such as more regular verb types than irregular verb types) may play a role in the overregulation. Therefore, because children come across many more regular verbs than irregular verbs, they overapply the regular rule.

On the other hand, it has also been argued that overregulation occurs because children make use of a grammatical rule (i.e., add -ed to a verb stem), which acts as the default rule. Hence, in cases where no irregular form can be retrieved, the default rule is applied. In German, the regular participle (-t) has a smaller type frequency than the irregular participle (-n) but exhibits overregulation in the same way as the English regular past tense. From this point of view, overregulations are argued to show evidence for symbolic or rule-based representations.

### Overregulation in Other Linguistic Phenomena

Although the primary example used here is the past tense, evidence of overregulation can also be found in the development of plural formation. In this case, the regular plural marker in English, -s, is overapplied incorrectly to nouns that would normally undergo an
irregular plural inflection (e.g., *mouses). In addition, children are also known to produce incorrect comparative and superlative forms of adjectives. In English, the comparative form of an adjective is created by adding the suffix -er to the adjective (e.g., *higher), and the superlative form is created by the addition of the suffix -est (e.g., *highest). A common error in comparative formations occurs with the adjective good where children incorrectly produce *gooder as the comparative form. Likewise, *goodest may be produced as the superlative form. Finally, irregular inflections themselves may be used as stems allowing for concatenation with a regular affix (e.g., for the past tense, *camed; for plurals, *mices; and for superlatives, *bestest).

Theoretical Implications

Overregulations in language development have been argued to provide evidence for nativist views of language acquisition (generally, views that language-specific innate processes play a large role in language acquisition). This view is predicated on the finding that overgeneralizations are attested implying that children do not simply acquire language based on input alone, as adults do not normally produce these errors and therefore there are no such errors in the input. Non-nativists, on the other hand, would argue that there are models of language acquisition which use error-free input that can simulate the U-shaped curve representative of past tense acquisition.

Although overregulation may not occur very frequently in language development, it nevertheless provides valuable insight into the mechanisms and processes that children make use of during language acquisition.

Renita Silva

See also Language Development; Language Production, Agreement in; Syntactic Production, Agreement in; Word Learning

Further Readings


LANGUAGE PRODUCTION, AGREEMENT IN

Agreement refers to the correspondence of some formal feature (person, gender, number) between an agreement controller (e.g., the subject noun) and syntactically related words in the sentence (e.g., the verb). Being syntactic in essence, this phenomenon provides a privileged window on the processes involved in grammatical encoding. This entry summarizes two major lines of research on agreement that have been pursued in psycholinguistics. The first line explores the influence of semantic and phonological factors in agreement to shed light on the issue of modularity in language production. The second line addresses agreement deficits in the production of specific populations.

Semantic and Phonological Factors in Agreement Production

In the early 1990s, Kathryn Bock initiated what became a whole research program on agreement production. She showed that so-called attraction errors by which an element with agreement features interferes in the realization of agreement (e.g., *cabinets* in “The key to the cabinets are on the table; * = ungrammatical sentence) can be reproduced and boosted in the laboratory by asking participants to complete sentences under time constraints. Experimental research showed that interference is primarily guided by the syntactic features of the “attracting” element, independently of the representation of these features at the conceptual level (e.g., plurality) and at the morphophonological level (i.e., in the form of the word, like the final s on English nouns). Nevertheless, agreement turns out to be under the partial guidance of conceptual and morphophonological features of the agreement controller. For example, the grammar of English allows plural verbs to occur with grammatically singular but conceptually plural collective head nouns (e.g., the faculty are threatening to protest). Along these lines, experimental research on attraction errors
nouns ending in number or gender morphologically (e.g., in Italian, is significantly reduced when the subject expresses attraction in number and gender logical form of the controller also modulates attraction manifested of agreement features in the morphological nine in these languages). Research showed that the eventually (e.g., the table, which is grammatically feminine) strictly grammatical gender, not represented conceptually (e.g., the table, which is grammatically feminine in these languages). Research showed that the manifestation of agreement features in the morphological form of the controller also modulates attraction. For example, attraction in number and gender is significantly reduced when the subject expresses number or gender morphologically (e.g., in Italian, nouns ending in o are typically masculine) as compared to when it fails to express it (nouns ending in e may be masculine or feminine). Similarly, the presence of case markers (i.e., markers that indicate the grammatical function of a noun, e.g., as a subject or object) was also found to reduce error rates as compared to ambiguous case marking. More generally, morphophonological influences were found to vary cross-linguistically as a function of the validity of the morphophonological marker in the language.

Different interpretations of conceptual and morphophonological effects on agreement have been proposed. Gabriella Vigliocco and colleagues interpreted them within an interactive framework in which grammatical encoding is maximally influenced by nongrammatical information. In the marking and morphing model developed by Bock and colleagues, nongrammatical influences are restricted to two specific stages of agreement production: Conceptual influences arise at the marking stage of sentence production when features from the message level are transmitted to the syntax, whereas morphophonological influences arise subsequently at the level of morphing when features are transmitted to structurally controlled elements (like verbs) and morphophonologically specified. Finally, Julie Franck and colleagues proposed that both conceptual and morphophonological influences arise at a stage of lexical selection during which the controller’s features are being selected within the functional lexicon (i.e., the lexicon for grammatical morphemes like nominal or verbal inflections), whereas these features are copied at a stage operating on the basis of syntactic principles, following characteristics of modular systems.

**Agreement in Specific Populations**

Children produce agreement markers on nouns (such as the final s in English) before they produce verbal or adjectival agreement morphology. Initial productions (of nouns as well as agreeing particles like verbs or adjectives) are characterized by a tendency of the child to produce unmarked forms (singular, masculine). Nevertheless, from early on, children also produce occasional plural and feminine forms and this consistently in the context of plural or feminine agreement controllers. This systematicity attests that even though young children tend to simplify their productions, they do so in virtue of grammatical constraints in agreement and not by chance. Across languages, children with specific language impairment usually show difficulties with inflectional morphology, with an overuse of default agreement forms. Difficulties with agreement production are also commonly reported in children with developmental dyslexia. Nevertheless, evidence is accumulating in favor of the hypothesis that the observed impairment with agreement in these populations does not result from a deficit in the syntactic machinery per se, but rather of impoverished processing and memory systems involved operating at the interface with syntax in language production. (For example, these populations show optional production of agreement markers, production in consistent contexts, and preserved sensitivity to agreement in comprehension.)

Similar profiles of agreement deficits are found in agrammatic speech in individuals with Broca’s aphasia, commonly described as involving a general breakdown in the production of grammatical morphemes. However, recent studies based on a finer grained approach to syntactic production in these patients have shown that, across languages, agreement morphemes are often either intact or better preserved than other inflectional morphemes like tense. The tree pruning hypothesis accounts for the patterns of impairments in people with agrammatism...
by assuming that higher projections of the syntactic tree are pruned, following the hierarchical order of functional categories. Under this account, agreement production by individuals with Broca’s aphasia is conceived of as a grammar-based phenomenon but driven by processing or working memory limitations.

Julie Franck

See also Aphasia; Planning in Language Production; Production of Language; Syntactic Production, Agreement in

Further Readings


**LANGUAGE PRODUCTION, INCREMENTAL PROCESSING IN**

Language production involves the generation of several successive levels of representation. These representations map between the thought to be expressed and the motor commands that articulate it. Current models of language production are in broad agreement about the nature of these representations. Speaking begins with the construction of a conceptual representation of the meaning to be expressed. This representation triggers grammatical encoding processes, which select the appropriate words from the mental lexicon and generate a syntactic structure to fix their linear order. Phonological encoding processes then generate the abstract sound structure of the utterance prior to articulation.

Models of language production agree that speakers need not generate all levels of representation for an entire utterance before beginning to speak. They propose that fluent speech output is accomplished by incremental processing, so that the articulation of early parts of an utterance occurs in parallel with the planning of upcoming segments. However, exactly how processing at different levels is coordinated remains a matter of dispute. In particular, there is disagreement about how much of an utterance must be generated at a particular level of representation before processing at the next level can begin.

This entry will introduce different proposals about the degree of incrementality operating during speech production. It will discuss the relevant findings for processes involved in the generation of grammatical and phonological structure.

**Generating Grammatical Structure**

There is conflicting evidence about how much of an utterance must be grammatically encoded prior to articulation. Some theories claim that the verb of a sentence must be retrieved before the generation of syntactic structure can commence. Verbs control the structure of clauses and there is evidence that the clause operates as a planning unit during speech production. Planning pauses in speech have been shown to occur more frequently between clauses than clause internally. Speech error data have been used to claim that the words for a whole clause are retrieved prior to speech onset. Word exchange errors such as “put the drawer in the cloth” suggest that the exchanging words *drawer* and *cloth* are retrieved in parallel. However, it is also proposed that these effects could occur during conceptual, rather than grammatical, processing.

Experimental evidence exists for tightly incremental planning. In recent years, eye-tracking studies have recorded the gaze patterns of speakers while they describe pictured scenes. These experiments demonstrate that speakers almost always fixate on pictured objects in their order of mention and rarely look ahead at objects to be named later (although some peripheral processing of immediately adjacent objects can occur). The time spent looking at an object is a function of the ease with which it can be identified, as well as of the ease with which the phonological form of its name can be retrieved. These findings suggest that we plan speech word-by-word with planning progressing only slightly ahead
of articulation. However, these experiments usually involved the production of one sentence structure to one fixed pattern of pictures. It is possible, therefore, that the observed visual and linguistic processing patterns are strategic rather than typical.

Evidence from reaction time experiments in which speakers produce more varied utterances is consistent with the first phrase as a planning unit. Sentence onset latencies (i.e., the time it takes to begin sentences) have been shown to increase as the complexity of the sentence initial phrase increases. In contrast, priming the retrieval of the verb in subject-verb sentences does not speed sentence production. For example, exposure to a verb (e.g., brushing) does not speed the production onset of a sentence with a semantically related verb (e.g., the woman is combing her hair). This finding suggests that the verb need not be accessed before speech onset.

It is possible that there is some flexibility in the degree of incrementality that speakers can adopt. However, there are also minimum planning units that must be employed due to grammatical dependencies. In many languages, grammatical agreement must occur between the elements of a sentence and the lower limit of advance planning will differ depending on the scope of the dependencies in a given language. For example, in many languages, the form of a determiner (e.g., the) is dependent on the gender of the noun it refers to (e.g., das rote auto, the red car).

Generating Phonological Structure

Generating the sound form of an utterance does not simply involve the retrieval of the stored phonological representation for each word, because the exact sound structure of a word is dependent on the context in which it is produced. Willem Levelt argued that prior to articulation, we construct a complete phonological word—a prosodic unit with one stressed syllable that can comprise several words. For example, the five words in the utterance, I gave it to her, can form one phonological word, l-ja-vi-ter.

A number of experiments provide support for this hypothesis. The time to begin to produce a word has been shown to increase with its number of syllables, suggesting that all syllables of a word are encoded prior to articulation. Moreover, sentence onset latency has been shown to be a function of the length of the initial phonological word, even when that unit comprises more than one lexical item (e.g., I-like-to).

However, the syllable latency effect has proved difficult to replicate. Moreover, it has been demonstrated using the picture-word interference task that priming may be limited to the initial syllable of disyllabic words in some two-word utterances. For example, naming a picture of a colored object (e.g., purple ball) is speeded by hearing the first syllable of the adjective (i.e., pur) but not the second syllable (i.e., ple). It is possible, therefore, that although the phonological word may be the preferred unit of phonological encoding, it may not constitute the minimal unit of phonological encoding.

As with grammatical encoding, it is possible that the degree of incrementality employed by speakers may vary across different speaking contexts. Similarly, there remains the issue of cross-linguistic differences in the scope across which dependencies operate during the generation of phonological structure.

Linda R. Wheeldon

See also Language Production, Agreement in; Production of Language; Prosody in Production

Further Readings

constraint that the architecture should exhibit *dissociations*, such that the lower levels still operate, and exhibit some sort of behavioral competence, in the absence (through damage or removal) of the higher layers but *not* vice versa. A substantial body of the neuroscience literature can be interpreted as demonstrating layered control systems in the vertebrate brain; layering is also an important theme in the design of artificial control systems, for instance, for autonomous robots.

**Layered Control in the Vertebrate Brain**

In many ways, the notion of layering is a common, often unspoken, assumption in neuroscience; however, the implications of the layered nature of the brain are not always acknowledged in a field often dominated by the study of the mammalian cortex. The idea of the brain as a layered architecture can be traced to the views of John Hughlings Jackson, a 19th-century neurologist inspired by the Darwinian revolution. According to Jackson, the various anatomical levels of the brain implement multiple functional levels of sensorimotor competence with higher centers supporting the same sort of “sensorimotor coordinations” as those below, but in a more complex fashion. This view contrasted with that of many of Jackson’s contemporaries for whom the highest levels of the brain, particularly the frontal lobes, were considered to be the seats of understanding and consciousness. Jackson strongly asserted that, although the frontal lobes may play a role in these more refined functions, the brain is a product of evolution and, therefore, all of its levels must be involved, in some way, with the coordination of sensing with action. Jackson also popularized the idea of *dissociation*; he argued that a breakdown at a higher layer should cause a reversion to the next highest layer of control. Discoveries of such dissociations between the vertical layers of the vertebrate nervous system were among the first findings of neuroscience. For instance, removing the cortex from a cat or a rat eliminates many major sensory, motor, and cognitive centers but leaves intact the ability to generate basic motivated behavior. That is, the animal still shivers when cold, escapes or fights when attacked, searches for food and drink, stops eating when sated, and so on. When most of the forebrain is removed, integrated behaviors can no longer be generated, but the capacity for elementary actions (walking, grooming, eating, etc.) is spared. With all but the hindbrain and spinal cord removed, the animal cannot coordinate the movements required for these actions; however, most of the component movements that make up the actions are still possible. The notion of a layered architecture is now being mapped out in more detail in the context of specific types of behavior. For instance, as illustrated in Figure 1, the vertebrate defense system—the control system that protects the body from physical harm—can be viewed as being instantiated in multiple layers from the spinal cord (reflexes), through the hindbrain (potentiates reflexes), midbrain (coordinated responses to species-specific stimuli), forebrain (coordinated responses to conditioned stimuli), and cortex (modification of responses according to context). In this system, the higher layers generally operate by modulating (suppressing, potentiating, or modifying) responses generated by the lower layers.

**Layered Control in Robots**

Although many control architectures for artificial agents include elements of layered control, the *subsumption architecture*, proposed by Rodney Brooks, illustrates the idea in perhaps its purest form. The principle of the subsumption architecture is that control systems are built incrementally, one layer at a time, with each new layer fully tested and debugged before another is added (emulating, perhaps, the process of natural evolution). New layers operate primarily by injecting their signals into lower layers thereby “subsuming” the lower level functionality. A key idea is that the system as whole does not construct integrated representations of the world; rather, sensory signals are processed differently at each level to implement relatively direct and behavior-specific mappings between sense data and the motor signals required to control the robot’s actuators. Robots built according to the principles of the subsumption architecture can operate in cluttered real-world settings, despite having relatively simple control systems, and show rapid responsiveness together with robustness to breakdown or damage (if higher levels become inoperable, the lower level system should still function). Arguably, however, the principles of subsumption have failed to scale to allow the control of robots with many actuators and multiple goals, that is, where the system is faced with particularly difficult problems of action selection.
Other Dimensions of Brain Organization

The nervous system is an intricately complex structure. Thus, although the vertebrate brain shows clear evidence of layered control, there will likely be other important governing principles in its organization. Indeed, a system that works by the principles of layered control alone may be too rigid to exhibit the intelligent, flexible behavior that vertebrate animals are clearly capable of (and currently far exceeds that achievable by robots). One proposal is of a centralized, or centrencephalic, organizing principle whereby a group of central, subcortical brain structures serves to coordinate and integrate the activity of both higher level and lower level neural systems. Although the notion of centrencephalic organization appears to oppose that of layered control, it is possible that the brain, although fundamentally layered, uses central integrative systems to manage and coordinate activity across the different levels. A better understanding of this full hybrid control architecture remains an important goal for the sciences of mind.

Tony J. Prescott

See also Common Coding; Decision Making, Neural Underpinnings; Natural Action Selection, Modeling

Further Readings

**Learning Styles**

Theories of learning styles suggest that learning varies across individuals, based on how the person perceives material (e.g., visually vs. auditorily) or how the person thinks about it (e.g., trying to solve a problem sequentially vs. trying to think about it globally). This entry describes what learning styles theories predict about human learning and discusses the lack of evidence supporting these predictions.

**Definition of Learning Styles**

The key prediction of learning styles theories is that learning and thinking ought to be faster and more effective if a person thinks about the material in a way that matches his or her style than if there is a mismatch. This preference is predicted to be stable and thus observable across time and with different learning tasks.

A few particulars of this definition require elaboration. First, learning styles are specific. The claim is not that “everyone learns differently” but that there is a limited number of styles. Second, learning styles reflect performance, not just preference. A sequential learner not only likes doing tasks sequentially rather than globally, he or she also performs tasks better that way. Third, learning styles occur on a continuum. It is usually assumed that a few people will show extreme preferences but that most will be somewhere in the middle. Fourth, learning styles are distinct from abilities. It is always better to have more ability than less, but styles are meant to be value neutral. There is nothing inherently better about being a visual learner or an auditory learner.

Different researchers have proposed different bases for learning styles. Some have suggested that a style is a relatively fixed part of one’s cognitive makeup, built into the mind. Others have suggested that learning styles actually reflect more broadly based personality types; the theoretical distinctions are meant to capture not only learning but also how one interacts with other people, for example. Still other theorists have suggested that learning styles are somewhat open to change, based on experience and on the moment-to-moment demands of the environment.

**Examples of Learning Styles Theories**

Testable theories of learning styles began to be proposed in earnest in the 1940s and 1950s. Since that time, dozens of theories have been proposed, many of which are rather complex. Some that may be described simply are shown in Table 1.

Researchers have developed testing instruments of reasonable reliability for a few theories, but for most

<table>
<thead>
<tr>
<th>Species</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Visual/auditory/kinesthetic</td>
<td>Learning is more effective if the material is perceived in the preferred modality</td>
</tr>
<tr>
<td>Converging/diverging</td>
<td>A preference for deductive thinking or for broad, associational thinking</td>
</tr>
<tr>
<td>Serialist/holist</td>
<td>A tendency to work through problems incrementally versus globally</td>
</tr>
<tr>
<td>Verbalize/visualize</td>
<td>A tendency to use words or to use mental pictures when thinking and reasoning</td>
</tr>
<tr>
<td>Field dependent/field independent</td>
<td>A tendency to see and think about related objects as a whole, versus separating an object from its surrounding context</td>
</tr>
</tbody>
</table>
Theories, there is not a reliable way to evaluate an individual person's style. This failing renders the theory untestable because, if one cannot say with certainty what a person's style is, there is no way of testing whether the style influences how that person thinks and learns. There are reliable methods of classifying learning styles for a few theories, but for these, validity is a problem. In other words, people with one learning style or another do not think and learn as the theory predicts they should.

Rather than thinking or learning in accordance with a particular style, individuals seem much more flexible, adapting their approach based on the demands of the task. Learning styles theories have been tested in a variety of situations, for example, how children learn basic print concepts, how college undergraduates study and understand complex material, and how children with learning disabilities learn to read. Learning styles theories are unsupported across these different people and tasks.

Despite the lack of scientific support, learning styles theories are widely believed to be accurate by educators, business people, and the general public. The bedrock of the idea is certainly optimistic. Educators hope to help struggling students, and business people hope to maximize efficiency. It would be a huge step forward if these goals could be better met through relatively minor changes in how information is presented or how people are encouraged to think about it. Unfortunately, the theory describing individual styles that could provide the basis for such changes has remained elusive.

Daniel T. Willingham

See also Metacognition and Education; Multiple Intelligences Theory

Further Readings


Legal Reasoning, Psychological Perspectives

Legal reasoning typically refers to the reasoning of judges deciding cases in trial or appellate courts. (Trial court judges decide cases; appellate court judges decide cases and more general legal or constitutional issues.) This entry describes the two main types of legal reasoning, the debates among legal scholars about the nature of legal reasoning, and the major differences between legal reasoning and scientific reasoning.

Legal reasoning includes both deductive and analogical (case-based) reasoning. In deductive reasoning, the decision maker is presented with a set of facts, searches statutes and legal precedents to discover the law that covers these facts, and reaches a verdict according to the law. The reasoning is syllogistic, with the law the major premise, the facts the minor premise, and the verdict the conclusion. In analogical reasoning, the decision maker examines the similarities and differences between the current case and earlier related cases and chooses the verdict that corresponds to the holdings of the cases it most resembles. Such case-based reasoning requires the ability to identify relevant prior cases, to discern their factual and legal similarities and dissimilarities to the present case, and to recognize which similarities and dissimilarities are relevant (e.g., the defendant’s state of mind) and which are not (e.g., the defendant’s name).

In practice, both methods require the resolution of various ambiguities. Usually the judge is not given a single set of facts, but two contradictory accounts, each suggesting the inevitability of a different legal outcome. In applying deductive reasoning, the judge may find several statutes or precedents that could be relevant, and the legal language may suggest more than one interpretation. In applying case-based reasoning, the judge must define the universe of possibly applicable cases and decide which are most like the current case and which, although apparently similar, are actually irrelevant. In both kinds of reasoning, the significance of a particular fact depends on its legal significance, and the significance of a particular law or previous case depends on the particular fact pattern of the current case.
Legal Formalism and Legal Realism

*Legal formalism* was a form of deductive legal reasoning promulgated by Christopher Columbus Langdell, who became dean of the Harvard Law School in 1870 and who transformed legal education from apprenticeship to professional education. According to formalism, the law is hierarchically organized with a few highly abstract principles at the top, from which a larger number of mid-level rules are derived, and finally a very large number of specific legal rules and case precedents, rather like the taxonomic system of Linnaeus (phyla, genera, species). Like the explorer who discovers a new species, the judge confronted with a new case could find its exact place in the ordered structure by comparing it with the low-level exemplars. Langdell promoted law as a logical, deductive science.

Formalism was sharply criticized by Oliver Wendell Holmes Jr. and later critics such as Roscoe Pound, Karl Llewellyn, and Benjamin Cardozo, who argued that application of legal rules does not yield definite answers in any but the easiest cases. Judges do not *discover* the defining distinction between one case and another by logical analysis; the boundaries are often fuzzy and overlapping, and what judges do is *create* the defining distinction. The fundamental principles and legal rules are important and provide guidance, but in many cases, are insufficient to determine the outcome.

In the 20th century, these criticisms were further developed in the *legal realism* movement. Its adherents were an eclectic group of political activists, admirers of the rapidly growing social sciences, and legal scholars who believed that the purpose of the law was as important as the letter of the law. The unifying theme was a rejection of formalism and a belief that legal doctrine played a limited role in actual legal decision making. They focused on the political, social, cultural, and psychological forces that influence legal decisions and on how legal decision makers actually reason, rather than how they justify their decisions. Legal realism has largely been absorbed into legal education, and consideration of the social context, purposes, and policy implications of the law is common. Few people still believe in strict Langdellian formalism, although many law courses are a blend of formalism and the considerations raised by the legal realists, and judicial opinions are often written in formalist language, describing the decision as unequivocally constrained by the law.

**Differences Between Legal Reasoning and Scientific Reasoning**

Scientists are more concerned with discovering valid generalizations than with specific cases. Legal reasoning is more like early psychiatry, in that individual cases are central. Excellence in legal reasoning involves finding factors that distinguish among apparently similar cases and general rules that unite apparently disparate cases.

When scientists are trying to decide among multiple competing hypotheses, they can design and carry out empirical research to generate new information. Judges do not have this power. Judges must work with the information given to them, and that information consists of what other people have said (attorneys, witnesses) or written (attorneys, judges, legal scholars) and their own background knowledge.

Scientists can also postpone making any decision at all. They can look at the evidence available and decide that it is inconclusive. Trial court judges have to make a decision on the basis of the evidence before them, even if it is ambiguous. Usually they have to decide for one party or the other. Appellate court judges have to make a decision based on the briefs and oral arguments. Their decisions have precedential force, defining the law for years to come, whether the case was close or clear. A 5-to-4 decision from the Supreme Court is as final as a unanimous decision. In science, no decision is reached when scientists are closely divided on an issue. Judicial decisions are always supposed to be final; scientific decisions never are.

Scientists think of variables like sanity or parental fitness or maturity as matters of degree, but legal decision makers are forced to draw bright lines and make categorical decisions: A 17-year-old is a child; an 18-year-old is an adult.

Ideas of free will and free choice are fundamental to legal thinking, and this creates a tension between law and the social sciences, which are, to varying degrees, deterministic. The judge’s task is to assign responsibility, which implies the assumption that people are personally accountable for their actions.
Some conditions can reduce responsibility, but free will is the starting point.

Finally, just as methodological rigor is central to science, adherence to the rules of law is central to legal decision making. Every decision that has precedential value must be justified by explicit discussion of the applicable law. Whether or not the rules actually determine the outcome, the judge must always provide a legal reason for the decision.

Phoebe C. Ellsworth

See also Analogical Mapping and Reasoning; Case-Based Reasoning, Computational Perspectives; Deductive Reasoning; Scientific Reasoning

Further Readings


Lie Detection

Most educated people understand that the link between body and mind makes plausible the physiological detection of deception. However, in the absence of a physiological response unique to lying, developing a valid lie detector has been both challenging and controversial. Modern lie detection relies on polygraph tests—collections of interrogation techniques assisted by physiological recording that are intended to detect criminal offenders, screen out dishonest job applicants, and identify personnel who pose security risks. Despite the widespread use of polygraphy in the United States, mostly by government agencies, the scientific community remains skeptical that polygraphy could have the high accuracy claimed by polygraph practitioners.

Relevant-Irrelevant Technique

Modern polygraph testing is computerized and involves obtaining digital representations of autonomic nervous system activity associated with palmar sweating, respiration, and blood pressure while individuals are asked different types of questions. The first broadly used polygraph techniques became known as relevant-irrelevant tests (RIT). The RIT includes a relevant or “did you do it” question (e.g., in the case of a woman’s rape, “Did you place your finger in Glenda’s vagina?”). The physiological reaction to relevant questions is compared to the response to truthfully answered irrelevant questions dealing with simple facts (e.g., “Is your name John?”). Guilt is inferred if the relevant question elicits the stronger response. The problem with the RIT is that the accusatory relevant question can be emotionally arousing even when answered truthfully, thus ensuring that it will elicit a strong physiological reaction, often much stronger than the response to the innocuous irrelevant question. Recognized for this bias against the innocent, the RIT has been largely replaced by the control or comparison question test (CQT).

Control (Comparison) Question Technique

The CQT, first introduced in the 1940s, is typically used in forensic applications. Like the RIT, the CQT includes relevant questions. However, the reactions to the relevant questions are compared to those from “control” questions that are thematically related to the content to the relevant question. Control questions, which are answered “no,” are believed to tap a “probable lie” based on the expectation that people typically engage in misbehaviors that are covered by the question. An example of a control question appropriate for an interrogation concerning a sex crime would be “Have you ever committed a sex act you were ashamed of?” CQT theory assumes that innocent individuals will be disturbed more by their denial to the control than to the relevant question because only the control question is likely to elicit a lie. By contrast, guilty individuals are expected to respond more to the relevant question because, for them, this question elicits a more important lie.

Regrettably, for the innocent, the significance of the accusation contained in the relevant and control questions is not equivalent. Only the relevant question deals with the criminal allegation, and only the response to this question can lead to consequences such as prosecution or public embarrassment. Just as is the case for the RIT, the CQT is biased against the innocent because only the relevant question deals with a consequential allegation. Hence, the accuracy
of failed CQTs is suspect. But can passed tests be trusted? Unfortunately, research has shown that liars can augment their physiological reactions to the control questions by lightly biting their tongue or doing stressful mental exercises, thereby beating a CQT. Neither failed nor passed CQTs can be trusted.

**Screening Techniques**

The most common applications of polygraphy occur with screening of law enforcement job applicants and of government employees with security clearances. Applicant screening tests deal with the integrity of prospective employees by, for instance, asking if they told the truth in response to hundreds of questions on their job application dealing with their trustworthiness on previous jobs, use of drugs and alcohol, meeting financial obligations, and so forth. Employees who already have jobs can be queried regarding their adherence to protocols designed to protect classified information entrusted to them. Screening test formats vary across settings, but many of the tests resemble an RIT with many relevant questions. Any relevant questions eliciting stronger responses than others are likely to become a basis for interrogation by the polygraph operator. If the examinee cannot explain the reactions to the satisfaction of the examiner, the test is failed. Although accurate statistics regarding the rate at which these tests are failed is lacking, anecdotal evidence indicates that preemployment tests are somewhat frequently failed whereas postemployment tests are rarely failed. This pattern is generally believed to reflect the examiner’s understanding that the cost of erroneously failing a qualified but untrained prospective employee is considerably less than that associated with failing a highly trained and thoroughly vetted current employee.

**Evaluation of the Scientific Foundation for Lie Detection Techniques**

Proponents of polygraph tests typically claim near infallibility for their techniques, especially for the CQT. Scientists at arm’s length from the polygraph profession have repeatedly evaluated the evidence in support of these claims and have concluded consistently that they are unfounded. The most recent and thoroughgoing evaluation of polygraph testing was carried out by a panel of more than a dozen scientists for the National Academy of Sciences. The panel held public hearings, had access to classified government documents and data on polygraph testing, and systematically evaluated the world literature on test validity. They concluded that the weak theory underlying polygraph tests renders implausible high accuracy claims, the generally poor quality of research on polygraph testing leads to overestimates of accuracy, the precise accuracy of polygraph tests is indeterminate, and there is no evidence that polygraph tests provide information about truthfulness that cannot be achieved by other methods.

An important question becomes why polygraph tests are used if they have such a weak scientific foundation. Polygraph tests are typically administered under circumstances where having the truth is important but difficult to obtain. Examiners are skilled interrogators adept at using the test occasion to leverage information from the examinee. Under the circumstances, important admissions and even confessions to otherwise unsolvable crimes are obtained. That anecdotal evidence indicates that such discovery frequently occurs speaks to the utility, not the accuracy, of polygraph tests. Indeed, the fastest growing application of polygraph testing involves their use with convicted sex offenders whose treatment progress and compliance with rules governing their release are difficult to monitor. Challenged to be truthful during their polygraph tests, sex offenders often divulge information about deviant sexual behavior and fantasies that monitoring programs use to help manage their rehabilitation.

**Probing Memory: The Guilty (Concealed) Knowledge Test**

Although conventional lie detection methods have, at best, a weak scientific foundation, there are memory-based detection methods that have a strong scientific rationale. The best known of these, introduced by David Lykken in 1959, is called the guilty knowledge test (GKT; sometimes referred to as the concealed information test). It is well established that a physiological response accompanies brain recognition of personally important information. The GKT is based on the principle that the perpetrator of a crime can be detected if bodily responses are recorded while being presented with crime-relevant memorial information that only the guilty and the police possess. For instance, in a case of rape, GKT questions such as the following might be asked while recording a suspect’s physiological activity:
“If you sexually assaulted Glenda, then you would know what weapon was used to force her compliance. Was it (a) a box cutter, (b) a baseball bat, (c) a hammer, (d) an ice pick? . . . In what room in her house did the crime take place? Was it (a) the pantry, (b) the attic, (c) the bathroom, (d) the garage?” Knowing that he grabbed an ice pick off the tool bench and assaulted her in the garage, the rapist would be expected to give the strongest response to these relevant items. An innocent suspect, not knowing the correct answers, would respond randomly to the multiple choice options. If enough questions are asked, there is little likelihood of an innocent person failing a GKT by chance. Likewise, if a recognition response is evident to almost all the correct alternatives, the likelihood of guilt is high. As one might expect, given the sensible rationale underlying the GKT, research has shown its potential to be highly accurate, with errors identifying the innocent as guilty especially unlikely.

However, there are several reasons why the GKT is seldom used in real-world applications. Basic research is needed to determine what criminals remember from their crimes. Without such knowledge, it is difficult to know how much confidence to place in a passed GKT because it is possible that the items do not deal with crime facts the perpetrator remembers. In addition, GKT item development requires considerable investigative work to identify material likely to be readily recognized by the perpetrator that is not likely to be known to other suspects in the case. Because the CQT, the main polygraph technique used by law enforcement, does not have these limitations, is believed by polygraph proponents to be very accurate, and has proven utility, the police see little advantage to substituting the GKT for the CQT.

Looking to the Future
Historically, lie detection has employed autonomic nervous system measures from peripheral body sites. Advances in neuroscience methodology now make relatively straightforward the recording of activity in the brain using techniques such as the dense array electroencephalogram and functional magnetic resonance imaging. These methods offer promise for the future, but how well they improve on autonomic measures has not been established. More important, research with these measures depends on the use of questioning formats that are similar to those already in use and thus are vulnerable to the well-known criticisms already leveled at existing lie detection techniques.

Conclusion
In conclusion, notwithstanding more than a half century of experience with conventional lie detection techniques, there are no generally accepted methods for detecting lying or identifying liars. Despite long-standing criticism that polygraphic interrogation is without scientific foundation, polygraph testing in the United States remains commonplace, and government agencies administer tens of thousands of these tests every year. This common usage points to the utility of the polygraph, the belief that polygraphic interrogation provides an effective vehicle for obtaining information from criminals and untrustworthy personnel that likely would be undiscoverable otherwise. As the National Academy of Sciences report noted, unlike other fields of scientific inquiry, polygraph research has not progressed over time, thus failing to strengthen its scientific underpinnings in any significant manner. Hence, it is unlikely that polygraph testing will be substantially improved or gain acceptance in the scientific community as valid. Memory-based methods, such as the GKT, that have sound scientific underpinnings, may profit from further refinement and appraisal.

William G. Iacono

See also Deception, Linguistic Cues to; Emotion, Psychophysiology of; Eyewitness Memory

Further Readings
Love

Love has the potential to make people both very happy and very unhappy. Yet, there seems to be no easy and unambiguous answer to questions as to what love is and how people fall in love. This entry will address some major areas of psychological research on love. First, taxonomies of love will be introduced. Then, the biological foundations of love will be explored. Finally, applications of the findings will be considered.

Theories

Taxonomies

Taxonomies are used to try to shed light on the different styles and kinds of love that may exist.

Romantic Love Styles

John Alan Lee proposed that there are three primary styles of love—eros, ludus, and storge—and three secondary styles of love that result from mixtures of the three primary styles—pragma, mania, and agape. Eros is an erotic kind of love that comes with strong passionate emotions and physical attraction. Ludus is a game-playing love that is uncommitted and tends to realize itself with a variety of partners. Storge is a friendship kind of love that does not come with emotions as strong as those of eros; in contrast to eros, it is relatively calm and unobtrusive. Pragma is a kind of calculating love that sees the partner in terms of attributes that are desired (or not desired) in a relationship. Mania is a highly emotional secondary style of love that alternates between euphoria and desperation or even agony. The third secondary love style is agape, which is a kind of communal and altruistic love that is very giving and compassionate but that usually does not appear in a pure form in romantic relationships.

Susan Hendrick and Clyde Hendrick used these love styles as the basis for their research program and suggested that those six love styles can be depicted in a six-dimensional matrix in which every person gets assigned a certain point on all of the six love styles to describe the “amount” of each love style. These styles are largely independent of each other. People can be especially high on one style or moderately high on several of them. Also, it is possible to experience different love styles with different partners. The love styles, therefore, are dependent not only on the individual but also on the partner, as well as on demographic factors like age, life stage, and so forth.

The Duplex Theory of Love

The duplex theory of love, developed by Robert J. Sternberg, has two parts. One part specifies the structure of love, the other part, how this structure comes to be. The two parts are called the triangular subtheory and the subtheory of love as a story.

The Triangular Subtheory of Love. The triangular theory of love holds that love can be understood in terms of three components that together can be viewed as forming the vertices of a triangle. These three components are intimacy, passion, and decision/commitment. Intimacy refers to feelings of closeness, connectedness, and bondedness in loving relationships. It thus includes within its purview those feelings that give rise to the experience of warmth in a loving relationship. Passion refers to the drives that lead to romance, physical attraction, sexual consummation, and related phenomena in loving relationships. Decision/commitment refers, in the short term, to the decision that one loves a certain other and, in the long term, to one’s commitment to maintain that love.

The three components of love generate eight possible limiting cases when considered in combination. Each of these cases gives rise to a different kind of love. It is important to realize that these kinds of love are limiting cases: No relationship is likely to be a pure case of any of them.

Nonlove refers simply to the absence of all three components of love. Liking results when one experiences only the intimacy component of love in the absence of the passion and decision/commitment components. Infatuated love results from
the experiencing of the passion component in the absence of the other components of love. Empty love emanates from the decision that one loves another and is committed to that love in the absence of both the intimacy and passion components of love. Romantic love derives from a combination of the intimacy and passion components. Companionate love derives from a combination of the intimacy and decision/commitment components of love. Fatuous love results from the combination of the passion and decision/commitment components in the absence of the intimacy component. Consummate love, or complete love, results from the full combination of all three components. Most loves are “impure” examples of these various kinds: They partake of all three vertices of the triangle, but in different amounts.

The Subtheory of Love as a Story. The kind of love triangles discussed in the previous section emanate from stories. The interaction of our personal attributes with the environment—the latter of which we, in part, create—leads to the development of stories about love that we then seek to fulfill, to the extent possible, in our lives. Various potential partners fit these stories to greater or lesser degrees, and we are more likely to succeed in close relationships with people whose stories more closely match our own. Although the stories we create are our own, they draw on our experience of living in the world. There is a potentially infinite number of stories, and the stories may contain some overlap.

Examples of stories are

1. Addiction. Strong anxious attachment; clinging behavior; anxiety at thought of losing partner.
2. Art. Love of partner for physical attractiveness; importance to person of partner’s always looking good.
3. Business. Relationships as business propositions; money is power; partners in close relationships as business partners.

The most common conception is of love as a travel story, or a journey that two people take together, trying to stay on the same path.

We may have multiple stories represented hierarchically, so that the stories are likely to vary in salience for us. In other words, we will prefer some stories over others, so that we may find partners differentially satisfying as a function of the extent to which they match our more salient stories.

Prototype Theory

Eleanor Rosch suggested that there are many concepts in everyday life that can be best described by means of prototypes. Prototypes are members of a category that represent the essence and typical features of the category members in a particularly good way. Other members of a category may differ in that some of their features are more or less prototypical of the category than others.

Beverley Fehr used the prototype approach to examine people’s conceptions of love. She found that people regarded characteristics of companionate love, such as caring and respect, as more prototypical of love than characteristics of passionate love, which would include features such as passion and sexual desire. Similar to the prototypical features of love, people regarded friendship as more prototypical of love than passionate love. Studies show that couples who had a rather prototypical view of love also felt more love for their partner.

Biological Theories

Biological theories focus on biological and physiological processes as well as knowledge of evolution to explain psychological phenomena.

Love as a Decision Bias

Douglas Kenrick and his colleagues view love as a system of decision biases that evolved over time. Human beings, similar to members of other species, have to make many decisions over the course of their day. In the domain of love, potential questions are: Who is the best person to mate with, how do I attract potential mates, and how do I retain them over a longer period of time? People often do not make decisions on an objective basis but rather have an (often unconscious) inclination toward one or the other action alternative. Kenrick suggests that their decisions are biased because there are some inborn decision biases that have evolved over the course of human history and development. The decision biases take the form of if-then rules where “if” refers to a certain condition in the environment and “then” constitutes a response that is designed to adapt to the environment. For example, when men are looking for potential partners, they tend to pay relatively
more attention than women to physical features like beauty and young age, which indicate fertility.

Different social situations necessitate different decision biases, which interact in dynamic ways with each other. These decision biases are the basis of all human behavior, so that although there may be cultural variations due to ecological differences, human behavior cannot be seen as a completely blank slate because the decision biases provide a certain framework.

**Love as a Means to Commitment**

David Buss suggests that love is primarily a device to achieve commitment—a means that helps bind people together through better or worse. He starts his explanation with changes that occurred when humans started diverging from their primate ancestors: Women's ovulation was concealed, so that men could not recognize when would be the best time to have intercourse with, and impregnate, them. This is one of the reasons men and women started to be engaged in long-term bonding. Buss has suggested that love evolved as a means to help people stay committed to each other even in difficult circumstances, for example, when they get sick or meet some other, more attractive potential mate.

Research has shown, for example, that when people feel love for their partner, they can better suppress thoughts of alternative attractive mates.

**Love From an Attachment Point of View**

Phillip Shaver and his colleagues applied the conceptual framework of attachment theory to adults. They suggest that for every behavior that is exhibited in attachment relationships of young children to their caregivers, there is also a parallel behavior in adult relationships. Shaver and colleagues suggest that three attachment styles exist in adults, just as in children.

A secure attachment style leads to people being comfortable being close to their close others, without any great fear of being abandoned or others getting too close to them. An anxious-ambivalent attachment style leads people to cling to their loved ones and to be afraid of losing them. An avoidant attachment style leads people to avoid closeness to others and to become anxious once those others seek proximity.

In addition to the attachment system, there are two other behavioral systems that may play a role in romantic love—the caregiving system and the sexual system. The caregiving system is triggered by others’ expressions of needs and attachment. Its goal is to help others in need and to reduce their misery. The sexual system is activated by the presence of a potential attractive and fertile partner. Its goal is to engage in sexual intercourse and, ultimately, to pass on one's genes to the next generation.

**Applications**

A feature of research on love is that it has applications to people's everyday lives. For example, people can assess their own love triangles or love stories or attachment styles. Assessing such patterns can help people better understand how they love and also what they are looking for in a partner. Partners can also assess the extent to which they are looking for the same things in a relationship.

Individuals and partners can also use theories of love to enhance their relationships. For example, people who have an anxious-ambivalent style of loving may have difficulties in their relationships and may wish to work on this style—either on their own or, preferably, in the context of psychotherapy—if it is causing them problems in their life. People first need to understand what their issues are in relationships. Theories of love can help them find out what these issues are. Then they can decide whether they want to resolve them, and how.

**Summary**

Theories of love address questions such as what love is, how it develops, how it can be assessed, and how it can be enhanced. We have considered several approaches, including taxonomic, prototypic, and biological ones. These theories differ in both the assumptions and assertions they make; however, they have in common that they attempt to provide plausible and empirically supported accounts of the nature of love.

Because the theories deal with somewhat different aspects of the phenomenon of love, they are not necessarily mutually exclusive. There may be elements of many theories that, in combination, help us understand the mysterious nature of love.

Robert J. Sternberg and Karin Sternberg

See also Attraction; Jealousy; Relationships, Development of
Further Readings


MACHINE SPEECH RECOGNITION

Speech represents the most natural means of human communication. Machine speech recognition, often called automatic speech recognition, is the automatic process performed by machine or computer to transform a speech utterance into a text consisting of a string of words. The term machine aims at making the distinction between machine speech recognition and human speech recognition (human speech perception). Machine speech recognition is also different from machine speaker recognition, which is the automatic process performed by machine to identify a speaker or to verify the identity of a speaker based on his or her voice. After a brief overview of the general steps involved in machine speech recognition, this entry introduces two of the most prominent machine speech recognition techniques.

There are different techniques for machine speech recognition, but generally speaking, they consist of a common series of steps. First, the acoustic waves of pressure corresponding to the speech utterances are transformed into electric signals by a microphone. These electric signals are then transformed into a string of feature vectors, usually called acoustic feature vectors. The feature vectors are representations of the spectrum and energy of the speech signal over short periods. Then, the extracted string of acoustic feature vectors is matched against previously stored models of sentences, words, syllables, or phonemes. The text string of words that best matches the incoming string of acoustic feature vectors is presented at the output of the machine speech recognition. Based on the type of input utterances, machine speech recognition can be classified as isolated word recognition or continuous speech recognition. Based on the generality of the models, machine speech recognition can be classified as speaker dependent or speaker independent. Based on the size of the vocabulary, machine speech recognition can be classified as small vocabulary (up to 100 words), medium vocabulary (up to 1,000 words), or large vocabulary (up to hundreds of thousands of words). Applications of machine speech recognition include voice dialing (e.g., digit recognition), command and control, form filling (e.g., data entry), web search by voice, and dictation (e.g., speech-to-text word processing).

Dynamic Time Warping

One of the most successful early techniques for machine speech recognition is called dynamic time warping (DTW) and is based on a combination of template matching and dynamic programming. Dynamic programming is a mathematical optimization process of finding the best (optimal) decisions in a recursive manner. A string of acoustic feature vectors corresponding to the input test utterance is matched consecutively against each stored reference template of feature vectors corresponding to training utterances. The test string of vectors and the stored string of vectors corresponding to each reference template form a search grid on which DTW finds an optimum path. The test feature vectors are warped nonlinearly in time (compressed or expanded) with the feature vectors of the stored templates. A matching score or distance is then computed between the
test utterance and each stored reference template. The input test utterance is recognized to be the utterance corresponding to the stored reference template that provides the highest score or lowest distance to the test utterance. The first DTW approaches used isolated words to create templates. Later, this technique was extended to connected speech by creating sentence templates made of concatenated word templates.

**Hidden Markov Models**

The most prominent modern technique for machine speech recognition is called hidden Markov models (HMM). In this technique, each stored model (e.g., a word or a phoneme) is characterized by a set of model parameters that consists of a sequence of states, an initial state probability vector, a state transition probability matrix, and an observation probability density function corresponding to each state. The initial state probability vector defines the probability of each of the states to be the entering state of the model. The state transition probability matrix defines the transition probability between each state and all other states, including itself. The observation probability density function defines the multivariate probability distribution of the feature vectors. The observations consist of acoustic feature vectors, which most commonly are in the form of mel frequency cepstral coefficients. These coefficients represent the spectrum of speech using the cosine transform of the logarithm of the spectrum on the mel frequency scale (a perceptual scale of pitch). The HMM technique consists of two phases: training and recognition. The training phase focuses on estimating the model parameters given multiple training utterances (observation sequences). The recognition phase focuses on searching for the hidden, most likely sequence of words and usually employs one of two kinds of algorithms—the A* algorithm (stack decoder) or the Viterbi algorithm (based on dynamic programming). For a given test utterance, the HMM provides the hidden sequence of states within sentences, words, and phonemes and its corresponding likelihood. The recognition process involves acoustic models (which characterize how words and phonemes are acoustically pronounced) and language models (which characterize the rules of combining different words into sentences). Both types of models are created from large numbers of training utterances and sentences for a given language.

**Performance**

The dream of machines capable of recognizing speech attracted many researchers during the past 6 decades. Many commercial products today claim accuracies between 95% and 99% for large-vocabulary continuous speech recognition in clean conditions and when models were adapted to the voice of the speaker. Yet, in spite of intensive research efforts, the performance of machine speech recognition is far behind the performance of human speech recognition. Some research studies show that the word error rate of humans is still about an order of magnitude lower than that of machines performing the same speech recognition tasks. There are many factors that make machine speech recognition a difficult problem. One is represented by the large acoustic variability and pronunciation variability both within speakers and across speakers. Another factor is the linguistic complexity of the task, which is due to the fact that words can be combined to form an infinite number of different sentences. Yet another challenge is the deterioration of speech signal due to environmental acoustic noise or channel noise.

*Sorin Dusan*

See also: Cohort Model of Auditory Word Recognition; Speech Perception; Word Recognition, Auditory

**Further Readings**


**McCOLLOUGH Effect**

All of us have experienced visual adaptation in some form or another. Stare at a bright red light for half...
a minute, look away and you will see a green spot on your retina before it fades away quickly. You just adapted to the (red) light and, as a result, saw a (green) color that is opposite or complementary to the original. In general, persistent exposure to a stimulus causes the neural circuitry responsive to the stimulus to adapt. The adaptation results in afterimages or aftereffects, namely, percepts that are dissimilar from the actual stimulus. Aftereffects are generally negative insofar as the feature value perceived following the adaptation is opposite that of the adapter (the stimulus that causes the adaptation), as in the preceding example.

Celeste McCollough’s discovery of an aftereffect that was contingent on a feature of the adapting pattern radically altered the study of adaptation. The aftereffect, known as the McCollough effect, has since been extensively replicated in other laboratories. It arises from the alternating presentation of stripes (gratings) of perpendicular (orthogonal) orientations of complementary colors for a modest time period (5–10 minutes). Figure 1 illustrates the adapting stimuli, which consist of red, horizontal and green, vertical gratings. Subsequent to the adaptation, achromatic (i.e., black and white) horizontal gratings appear greenish and achromatic vertical gratings pinkish, opposite to the colors shown during adaptation. Thus, similar to the overwhelming majority of aftereffects, the McCollough effect is a negative aftereffect.

The McCollough effect is a dramatic departure from other negative aftereffects, however. In contrast to simple aftereffects that do not require a test stimulus for the misperception to be evident, the McCollough effect is a contingent aftereffect that is created by relatively brief experimentally induced correlations between stimuli that are usually uncorrelated in the real world. The aftereffect is different depending on whether the test stimulus is horizontal or vertical in the previous example. Also, in contrast to simple aftereffects that last a few seconds, the McCollough effect is stable for weeks, or even months.

There are at least two other fascinating aspects of the McCollough effect. The observer does not have to maintain fixation on a point on the adapter; in fact, the observer can let the eyes wander over the pattern, and the resulting aftereffect will be largely unaffected. Second, the effect does not transfer interocically: If adaptation is limited to one eye, no discernible negative aftereffect is observed in the non-adapted eye. Given that the McCollough effect is contingent on adapter orientation, that orientation selectivity does not emerge before primary visual cortex (or V1), and that orientation-tuned cells in V1 typically have binocular responses, one would expect the McCollough effect to transfer to the non-adapted eye. That it does not is surprising, and suggests a subcortical locus, where information from the two eyes is segregated.

Stemming from the McCollough effect, research has expanded into investigations into mechanisms or theories to explain the McCollough effect, explorations of aftereffects on other visual dimensions, and discoveries of positive contingent aftereffects.

**Figure 1  McCollough effect**

Note: A typical sequence of stimuli used to elicit the McCollough effect. To obtain an effect, the stimuli are not in grayscale, as shown, but colored.
offer a biologically plausible model. In this regard, McCollough originally proposed that cells that were sensitive to both color and orientation, which lie in visual cortex, adapted. A more recent and influential neural network model by Stephen Grossberg and colleagues accounts for the spectrum of empirical findings related to the McCollough effect and is based on known mechanisms in biological vision.

Contingent Aftereffects
In addition to inspiring models and theories of learning and visual function, the discovery of the McCollough effect spawned a veritable cottage industry of contingent aftereffects. A number of aftereffects that are contingent on color have been found by pairing color with visual dimensions such as movement and spatial frequency. Complementary to the McCollough effect, which is a color aftereffect contingent on orientation, a tilt aftereffect that is contingent on color has been found. Contingent aftereffects with visual dimensions other than color have also been obtained. For example, aftereffects that are contingent on the direction of stimulus motion have been reported in a number of studies. Motion aftereffects that are contingent on stimulus luminance, size, direction of gaze, binocular disparity, the orientation of a stationary grating superimposed on a moving textured surface, and even the color or pattern of a stationary area surrounding the moving stimulus have been reported. Negative aftereffects that are contingent jointly on two different dimensions of a stimulus have also been reported. Finally, a long-lasting contingent aftereffect has been observed in the auditory system.

Positive and Double Negative Aftereffects
Not all contingent aftereffects are negative. There have been reports of a positive aftereffect; that is, the test color perceived following the adaptation is the same as the color of the oriented adapter. In another study, observers simultaneously viewed alternating orthogonally oriented achromatic gratings in one eye and alternating red and green homogeneous textureless fields in the other eye. When the eye that adapted to the color fields viewed achromatic test gratings later, observers reported a classical negative aftereffect, but when the eye that adapted to the oriented gratings viewed the same gratings, observers reported a positive color aftereffect. More recently, a double negative aftereffect has been reported. Binocular or monocular exposure to alternating red and achromatic horizontal gratings for equal durations leads to an aftereffect of at least 24 hours duration in which test achromatic horizontal gratings appear reddish, that is, the color that is paired with the adapter orientation. The effect, termed the anti-Mccollough effect, transfers 100% to the non-adapted eye, suggesting a neural locus on a higher level than that of the classical McCollough effect. One interpretation of the finding is that the neurons in a lower order area adapted to the stimulus, and neurons in a higher order area downstream adapted to the signal from the lower order area, giving rise to a negative aftereffect of the classical negative aftereffect. The findings are in accord with the idea that signals from higher order areas have greater access to visual awareness than those from lower order areas. Finally, studies have used contingent aftereffects to distinguish fast, preconscious from slow, conscious color processing by demonstrating that McCollough aftereffects can be induced at frame rates at which conscious perception fails.

The discovery of the McCollough effect in 1965 revolutionized the area of sensory adaptation, but its inner workings still remain somewhat of a mystery. Nevertheless, contingent aftereffects, in general, can be promising tools for understanding how the brain perceives.

Bhavin R. Sheth

See also Face Perception; Reinforcement Learning, Psychological Perspectives; Stereopsis

Further Readings

Memory, Interference With

There are two types of memory interference: proactive interference and retroactive interference. Proactive interference occurs during memory retrieval when a later memory trace is hindered by...
a highly similar earlier memory trace. *Retroactive interference* can occur during both memory retrieval and memory consolidation (strengthening). It occurs during retrieval when an earlier memory trace is hindered by a highly similar, later memory trace. Retroactive interference with consolidation occurs when the consolidation of a recently acquired memory trace is hindered by further cognitive activity, drugs, brain lesions, or seizures. Memory interference theory differs from another theory of forgetting, decay theory, in that it attributes forgetting to interference rather than to the sole passage of time.

**Retroactive Interference**

*Retroactive Interference With Memory Consolidation*

Retroactive Interference by Further Cognitive Activity

The notion that forgetting might occur due to memory interference was first put forward in 1900 by the German psychologist Georg Müller and medical student Alfons Pilzecker. They noted during their early experimental work on healthy participants that recently learned material often seemed to “pop into consciousness” in an unwilled manner, especially when their participants were not engaged in any mentally effortful activity following new learning. Müller and Pilzecker hypothesized that this “perseveration” of recently learned material reflected a transiently continued activity of learning-related physiological processes that served the “consolidation,” that is, the strengthening of new memory traces.

They posited that the consolidation of a recently acquired memory trace could be hindered by subsequent cognitive activity; they called this effect “retroactive interference.” Their work showed that, indeed, more nonsense syllables were forgotten by their participants when the learning of such nonsense syllables was followed by the learning of further nonsense syllables (filled delay) than when it was followed by a period of rest (unfilled delay). This retroactive interference effect was also apparent when the interpolated material was not similar to the earlier material, that is, when the learning of nonsense syllables was followed by a picture description task; this result indicated that this interference effect was nonspecific. Further evidence for such nonspecific effects comes from John Jenkins and Karl Dallenbach, whose much-cited study demonstrated that more nonsense syllables were forgotten when the learning phase was followed by routine daily activities as opposed to a period of sleep.

In an attempt to explain how (nonspecific) cognitive activity might impede the consolidation of an earlier memory trace, the psychologist John Wixted argues that any cognitive activity engages the consolidation system to allow for the retention of this activity and its associated material. Moreover, he hypothesizes that consolidation resources are not unlimited. Therefore, according to Wixted, post-learning cognitive activity impedes the consolidation of an earlier memory trace because it actively deprives the earlier memory trace of limited consolidation resources.

However, the magnitude of this interference effect is not constant over time. Indeed, the most important and influential finding to have emerged from Müller and Pilzecker’s as well as Ernest Burton Skaggs’s early memory research is that of a temporally graded effect of further activity on early memory traces: Retroactive interference is largest when further activity occurs immediately following the learning of early material, and it decreases with increasing delay in the onset of further activity. This temporal gradient of retroactive interference indicates that memory traces are initially highly fragile and vulnerable to interference but that they consolidate and become less susceptible to interference over time.

**Increased Retroactive Interference by Further Cognitive Activity in Patients With Amnesia**

Recent neuropsychological work on amnesic patients by Nelson Cowan, Sergio Della Sala, Michaela Dewar, and their colleagues has shown substantially augmented effects of post-learning activity on recent memory traces in some patients with anterograde amnesia, indicating that these patients are especially vulnerable to retroactive interference. In line with the consolidation theory, the research has revealed a steep temporal gradient of retroactive interference in at least some amnesic patients. This gradient demonstrates that, even in amnesic patients, memory traces can consolidate over time, provided that this time is devoid of further activity.

**Retroactive Interference by Brain Lesions**

It is not only further cognitive activity that differentially interferes with older and newer memory traces. Around the same time as Müller and Pilzecker were investigating consolidation and retroactive interference in healthy participants, the French psychologist Théodule-Armand Ribot noted that certain types of pathological damage to the brain had
a more detrimental effect on recent pre-morbid than remote pre-morbid memories, implying that earlier memories are less vulnerable to pathological brain damage than are later memories. This effect, which has been widely reported and is known now as “temporally graded retrograde amnesia,” is in line with Müller and Pilzecker’s consolidation theory, in that it is strongly suggestive of a gradual strengthening of memories over time.

Retroactive Interference by Drugs

Pharmacological animal work has shown that memory consolidation can also be hindered by certain drugs such as protein synthesis inhibitors, which are toxins or antibiotics. This work usually involves the learning of a response such as shock avoidance, followed by the injection of a protein synthesis inhibitor, and subsequent memory testing. As with retroactive interference by further cognitive activity, protein synthesis inhibitors have a temporally graded interference effect on new memory traces. They are most detrimental to a new memory trace when injection occurs immediately following learning, and they become less effective with augmenting delay between learning and injection.

Retroactive Interference With Memory Retrieval

Interpolated cognitive activity not only interferes with memory consolidation of an earlier memory trace; it can also interfere with the retrieval of an earlier memory trace, provided that the memory trace resulting from the interpolated cognitive activity is highly similar to the earlier memory trace. Similarity can take various forms: Two or more memory traces can be similar in type, that is, visually, phonologically, or semantically, or they can be similar because of a mutual association with a retrieval cue. For example, a computer login screen might cue one’s private as well as one’s work computer password.

One of the classic early experimental psychology paradigms to demonstrate a similarity effect of retroactive interference was that by the psychologists John McGeoch and William MacDonald. They showed that the magnitude of forgetting of new verbal material was closely related to the degree of similarity between the early and interpolated later material. Forgetting of a list of adjectives was lowest following relatively dissimilar interpolated material (the reading of jokes, referred to as “rest” by the authors), and it increased progressively from interpolated three-figure numbers to interpolated nonsense syllables to interpolated antonyms and finally to interpolated synonyms.

Retrieval-based retroactive interference theory continues to dominate psychological memory interference research, and most current psychology textbook definitions of retroactive interference refer solely to retrieval, not consolidation.

The general consensus is that the attempted retrieval of a particular memory trace also activates those memory traces that are highly similar to the to-be-retrieved memory trace, or which are associated with the same retrieval cue. Such activation of multiple memory traces is said to lead to competition for retrieval and, thus, to interference. According to this theory, retroactive interference arises if the attempted retrieval of an early memory trace also activates a later acquired and currently stronger memory trace. For example, when asked to verify one’s previous phone number it is likely that one’s new phone number will be activated and indeed initially retrieved.

In the laboratory, such retrieval retroactive interference can be reliably produced via paired associate learning paradigms (AB-AC paradigms). These paradigms consist of the learning of a list of word pairs, each comprising a cue word and an associated word, for example, *dog-train*. List 1 learning is then followed by the learning of a second list of word pairs, which consist of the same cue words and new associated words, for example, *dog-sun*. This interpolated list tends to interfere during the subsequent cued recall of List 1 (i.e., *dog-*). It has still not been established conclusively whether the interpolated memory trace is (transiently) stronger than the older trace or whether it might actually weaken the old trace within memory storage, for example, as a result of the active inhibition of the old trace during learning of the new trace.

Proactive Interference

One of the first explicit references to proactive interference comes from the 19th-century German psychologist Hugo Münsterberg, who always carried his pocket watch in a particular pocket. He noticed that after changing pockets he continued to automatically reach to the old pocket watch location. Such instances of proactive interference are believed to occur because a cue (in this particular example the intention to check the time) is associated with both an old, established memory trace and a new memory trace. Given the greater strength of the
old memory trace, the cue results in the retrieval of the old memory trace. Similar situations can arise when one changes one’s computer password or one’s address, or when a woman marries and takes on her husband’s surname.

**Early Experimental Work on Proactive Interference**

In the 1950s, the psychologist Benton Underwood went as far as to propose that most forgetting could be accounted for by such proactive interference as opposed to by retroactive interference. He argued that the then-prominent retrieval proactive interference theory could not explain why some participants showed extensive forgetting of nonsense syllables over a period of 24 hours when this period was not filled with any similar material (i.e., when it was spent engaged in everyday activities at home). Having examined various studies, he noted that the amount of forgetting over 24 hours was in fact closely related to the number of study lists learned prior to the learning of the to-be-recalled list.

Importantly, however, this relationship was only present in studies in which the learning of prior lists was massed. The amount of forgetting was unrelated to the number of prior lists when these were spread over a longer time frame, as tends to be the case outside the laboratory. Proactive interference theory also appeared to be unable to explain the benefit of interpolated sleep that was demonstrated by Jenkins and Dallenbach. Underwood's claims that proactive interference was a major source of forgetting could thus not be sustained for very long.

**Increased Proactive Interference in Patients With Executive Impairment**

Even though no longer in the top list of forgetting causes, proactive interference does of course produce some forgetting as evinced in the Münsterberg example presented earlier. It is especially often observed in the laboratory in patients with damage to the executive system who have difficulties in the inhibition of irrelevant stimuli and distracters. In the aforementioned AB-AC paradigm, for example, dys-executive patients often continue to recall AB associations during the AC learning trials. These patients are also often reported to show increased levels of proactive interference with retrieval. It is hypothesized that such increased levels of experimentally induced retrieval interference are the result of the patients’ general impairment in the inhibition of irrelevant, in this case competing, stimuli.

*Micheala Dewar and Sergio Della Sala*

**See also** Amnesia; Memory, Neural Basis; Memory Recall, Dynamics

**Further Readings**


**MEMORY, NEURAL BASIS**

The ability to learn new information, and subsequently remember it, is critical to all animals. Whether it is apparent in response changes to repeated stimuli by an invertebrate, or by the collection of events in one’s personal past that shapes the complex behavior of a human, memory allows an organism to effectively adapt to the environment. The study of memory processes has been arguably the most exciting and productive area of behavioral neuroscience in the past 50 years. Because learning is ubiquitous across animals, it has been possible to study its biological mechanisms in animal models in a way not possible for most other human cognitive capacities. This work has allowed researchers to understand neural mechanisms of memory at multiple levels of analysis, including molecular interactions, plasticity in circuits, and the roles of different anatomical systems. In this entry, the critical contributions at each
of these levels of analysis will be discussed. Findings from human amnesic patients are described in terms of how these studies led to the idea that there are different brain systems that support different kinds of memory. In subsequent sections, studies at the circuit level are described which focus on how associations are formed in the brain. Finally, the entry outlines work examining the molecular mechanisms of long-term plasticity using brain slices.

**Memory and Neural Systems**

Most modern discussions of memory and the brain begin with the case of Henry Molaison, who was known as patient H. M. to most psychologists and neuroscientists until his death in 2008. Molaison had suffered from serious epileptic seizures starting in childhood, and he underwent bilateral surgical removal of the medial temporal lobe in 1953 in order to remove the epileptic foci. While this surgery was effective in that the frequency of seizures decreased substantially, it resulted in a severe impairment in memory. Molaison was virtually unable to retain any new information, such as the names of new people he met. What was surprising at the time was that this profound memory impairment occurred with a background of normal intellectual functioning. H. M. was able to perceive, reason, and use language relatively normally. Thus, it appeared that memory could be selectively affected by a lesion, demonstrating that a distinct neural system supports memory processing. Although memory impairment had been reported before as the result of brain injury or neurological disease, such as Alzheimer’s disease, these cases typically involved more widespread cognitive impairment. The careful study of patient Henry Molaison, and of other amnesic patients that followed, promoted the view that memory can be localized in the human brain.

The brain structures that are often damaged in amnesia include the hippocampus and cortical regions in the medial temporal lobe that surround this structure. This medial temporal lobe system is organized in a hierarchical fashion. The hippocampus receives and sends projections to the entorhinal cortex, which is interconnected with the parahippocampal and perirhinal cortices (Figure 1).

![Figure 1](image1.png)

**Figure 1** Magnetic resonance images of the human brain, showing regions that are damaged in medial temporal lobe amnesia

*Note:* The panels show frontal sections, with the top of the figure representing the top of the head, and the image being in the plane of the face. The top panel is more anterior (closer to the face) than the bottom panel. The hippocampus is shown in black on the right side in both images. The entorhinal cortex is shown in dark gray and the perirhinal cortex is shown in white in the anterior image. These structures do not extend to the posterior section. The parahippocampal cortex appears in more posterior sections and is shown in white in the lower panel. These brain structures can be seen as on the left side as they appear in the magnetic resonance image.
These regions are, in turn, interconnected with a wide range of high-level association cortical areas, including frontal, temporal, and parietal regions. With this architecture, the hippocampus is positioned to receive highly processed input from all over the brain. This is consistent with the idea that memory representations are complex and multimodal. Amnesia can also result from damage to structures in the diencephalon that are connected to the hippocampus via the fornix.

While much focus has been on the hippocampus, it is clear that the surrounding cortices make additional contributions above and beyond their role as conduits for information. While damage restricted to the hippocampus results in a significant memory impairment, a far more significant and profound impairment results when these cortical regions are damaged as well, as in the case of Henry Molaison. These results suggest that these regions play a role in memory processing, particularly in the primate brain in which these cortical areas are well developed and provide the major input into the hippocampus. A currently unresolved question is whether the hippocampus and these cortical areas play similar or slightly different roles in memory.

A second major insight gained from the study of amnesic patients is the fact that the eventual storage site of memories is not in the medial temporal lobe. Patients with medial temporal lobe damage are impaired at learning new information, but memories from the remote past are intact. Thus, Henry Molaison’s knowledge of facts about the world and vocabulary from before his surgery remained intact, although he was unable to acquire much in the way of new facts, and words that were invented after his surgery did not enter his vocabulary for the most part. These findings indicate that remote memories are stored in regions that are intact in amnesic patients. A likely site is the cerebral cortex. Cortical regions outside the medial temporal lobe, particularly the lateral and inferior temporal lobe, are likely candidates. These regions receive output from medial temporal lobe cortices, and enduring memory traces may be set up here. By another view, the medial temporal lobe structures remain important for retrieving memories about events, even those that were learned remotely. By this view, individual elements of memories may be stored in cortex, but medial temporal lobe structures are needed to bind together these traces to reexperience all elements of an episode.

Some support for the idea that remotely learned information is stored in temporal cortex comes from the study of patients with semantic dementia. Semantic dementia is a progressive neurological disease that involves degeneration of temporal lobe cortex. These patients lose knowledge of the world, and appear to have lost older memories, but new memories can still be formed, at least as long as medial temporal lobe regions are not affected by the disease.

Whereas remote memories are preserved even after extensive medial temporal lobe damage, more recently learned information can be vulnerable. Damage to the medial temporal lobe can result in amnesia for information learned before the injury, a phenomenon known as retrograde amnesia. In some cases, the extent of retrograde amnesia can be fairly brief (a few days), but in other cases, it may extend for many years. A primary factor in the duration of retrograde amnesia is the extent of damage to the medial temporal lobe memory system. Those patients with complete damage to the hippocampus may have difficulty remembering information for several weeks or months before their lesion, whereas patients with extensive damage to medial temporal lobe cortices as well as hippocampus may have retrograde amnesia extending throughout most of their adult lives.

The existence of retrograde amnesia indicates that, for some time after initial encoding, the medial temporal lobe is necessary for memories to be retrieved. This finding suggests that a kind of “consolidation” takes place, with memory traces changing gradually across time. These changes would lead to a gradual independence of these memory traces from the medial temporal lobe. One possibility is that memory traces are set up in the cortical areas, but that for an extended time, the medial temporal lobe is needed to access them. As noted earlier, it is also possible that the medial temporal lobe is always needed for retrieving detailed contextual memories.

The Declarative/Nondeclarative Distinction

In addition to the importance of the medial temporal lobe in memory functioning, another major lesson learned from the study of Henry Molaison is that there are multiple forms of memory that depend on different brain regions. Amnesic patients show impairments in memory for events and facts, known as declarative memory. Declarative memories share the property that one can “declare” them; that is, they are
verbalizable. Despite impaired declarative memory, amnesic patients are able to show learning in different domains. Skill learning has been one of the most widely studied forms of intact learning in amnesic patients. These patients are able to learn to perform new motor skills at the same rate as neurologically intact subjects. One of the visuomotor skill tasks commonly used in the laboratory to assess performance of amnesic patients is the rotor-pursuit task. In this task, subjects attempt to keep a stylus in contact with a rotating disk. At first, it is difficult to keep the stylus in contact, but it becomes easier with practice. Although an amnesic patient may have difficulty in remembering the testing episode from one day to the next, he or she would show absolutely normal performance as measured by the decrease in errors over trials.

Another visuomotor task in which amnesic patients show normal performance is the serial reaction time task. In this task, an asterisk prompt appears in one of four locations, and subjects press a key corresponding to the position as soon as it appears. The asterisks appear according to a complex sequence that is not readily apparent to the subjects. Nevertheless, subjects react more quickly to each prompt with practice. It also appears that they have learned the sequence, because if the asterisks begin to appear randomly, performance slows down. Amnesic patients show normal learning of this sequence despite their poor memory for the testing episode.

The motor skills tested in the laboratory are similar to the more complex skills learned in daily life in that what is learned is often difficult to verbalize. While most of us learn to ride a bicycle with practice, we generally have a great deal of difficulty telling someone how to do it. Another example is learning to drive a car with a manual transmission. One must learn by doing, and not through verbal instructions. This characteristic of skill learning sets it apart from the kind of learning that is impaired in amnesic patients that is dependent on medial temporal lobe structures. The distinction between our declarative memory for facts and events, which we can verbally describe and which we are aware of learning, and non-verbalizable knowledge that we acquire in a motor skill learning task is fundamental to differentiating between the roles of different brain systems in memory.

Another memory domain in which performance of amnesic patients is intact is perceptual priming. When people see a stimulus, such as a word or picture, they can process it a little more effectively (faster and more accurately) if they had been presented with it previously. It appears as if the initial presentation leaves some sort of trace that allows more efficient processing when it is presented again. This facilitation does not depend on remembering that the primed stimulus had been presented. In fact, amnesic patients who are unable to remember that stimuli had been presented at all nevertheless exhibit normal priming. This finding demonstrates that the neural changes that support priming are independent of the medial temporal lobe memory system.

Priming and motor skill learning share the property that neither requires conscious awareness of what has been learned. However, in other ways, these two kinds of learning are different. Skills are generally learned gradually and incrementally. Priming, on the other hand, occurs rapidly, even after a single exposure to a stimulus. These two forms of memory also depend on different brain systems. Motor skill learning depends on the striatum and its interaction with the cerebral cortex, primarily the frontal lobes. Patients with neurological diseases that affect the striatum, including Huntington’s disease and Parkinson’s disease, show impaired ability to learn new motor skills in addition to their difficulties in motor performance. Neuroimaging findings also show activation in the striatum that is related to motor skill learning.

Priming appears to involve changes in the sensory cortex that processes the primed stimuli. Most of the work in this area has been done in the visual modality. The second viewing of a stimulus is generally accompanied by less activity in the extrastriate visual cortex than the first viewing as measured by neuroimaging techniques such as functional magnetic resonance imaging. This finding is consistent with the idea that the initial exposure makes it easier to process subsequent exposures, and thus less blood flow is needed to the neural region representing the item. This efficiency may be realized by the representation becoming more tuned to the task at hand. For example, if one is identifying objects, the second time around, those elements of the representation that are key for identification will be more activated, and those elements that are not critical will be activated less, thus resulting in behavioral priming (faster identification) and neural priming (less blood flow to the region).

The most compelling evidence for separate memory systems comes from double dissociations in memory performance observed in neuropsychological
patients. Whereas amnesic patients exhibit normal skill learning and impaired declarative learning, the opposite pattern is observed in patients with basal ganglia dysfunction, such as those suffering from Parkinson's disease, who are able to remember training episodes but who exhibit impaired skill learning. A double dissociation has also been reported between declarative memory and perceptual priming. A few patients with lesions in extrastriate visual areas have been described who show deficits in perceptual priming despite normal declarative memory for the same items. These double dissociations are important in that they demonstrate that skill learning and priming are not simply easier than declarative learning, as one could argue based solely on the intact results from amnesic patients, but rather rely on independent brain systems.

Neural Circuits in Associative Memory

In the quest to describe how the brain contributes to memory processes, a particularly fruitful approach has been to study relatively simple forms of learning in animal models. For example, in Pavlovian conditioning paradigms, what is learned is an association between a conditional and an unconditional stimulus (CS and US). A tractable approach is to identify the neural circuits that process the CS and US and sites of their convergence. These convergence areas would be likely candidates for sites of plasticity and thus could be targeted for physiological investigations.

This approach has been applied to discover the neural substrates of Pavlovian conditioning of the eyeblink response in the rabbit. Rabbits have a nictitating membrane, or third eyelid, that moves in response to an unconditional stimulus, such as a puff of air to the cornea. If this unconditional stimulus is consistently paired with a neutral conditional stimulus, such as a tone, a conditional response will gradually develop in response to the tone alone. Early work showed that animals retained the conditional response even after the removal of the forebrain. These results strongly implicate the cerebellum, which receives sensory input from the brain stem and plays an important role in precisely timed behaviors. The conditioned eyeblink response is timed optimally so that the nictitating membrane is maximally closed when the unconditional response onset is expected. That is, if a rabbit is trained with the airpuff occurring 500 milliseconds after the onset of a tone, the rabbit's conditioned eyeblink response will peak at 500 milliseconds after the tone onset.

The main input to the cerebellum is from sensory nuclei in the brain stem. Through a series of lesion studies, the pathways that are necessary and sufficient for this learning have been mapped out. The lateral pontine nucleus sends information about auditory conditional stimuli to the cerebellum. Information about the unconditional stimulus is conveyed via the climbing fibers from the inferior olivary nucleus. Lesioning or inactivating this nucleus in a well-trained rabbit does not abolish the conditional response. Rather, the conditional response is gradually extinguished over trials. This outcome arises because the input of the unconditional response is removed from the learning circuit. It is as if the air puff is no longer being presented.

Input from the conditional and unconditional stimuli converge on cerebellar Purkinje cells. These are remarkable neurons with extensive arbors. Based on the anatomy of the cerebellum, David Marr developed a model in 1969 in which convergent inputs could become associated by impinging on the Purkinje cells with close temporal proximity. When both climbing fibers and the conditional stimulus inputs are activated, long-term changes occur in the synapses. Although the conditional stimulus inputs are initially weak, they are strengthened through pairing with the much stronger climbing fiber inputs. After training, presentation of the conditional stimulus alone becomes adequate to affect the Purkinje cells. In these neurons, the strengthened synapses are actually inhibitory, so these cells have reduced output in response to the conditional stimulus compared to the level before learning. This reduced output ultimately modulates motor regions of the brain, resulting in a conditional response. The immediate output of the Purkinje cells is the deep nuclei of the cerebellum. Input from the conditional and unconditional stimuli also converge in this region, so the deep nuclei may work together with the Purkinje cells in the cerebellar cortex to produce optimally timed learned responses.

By studying this simple form of learning, it has been possible to identify the essential site of the memory, or the “engram.” A similar approach has been taken to study the circuit supporting the conditioning of an emotional response. One procedure that has been studied in detail is conditional fear in the rat. As rats are prey animals, they have evolved
robust neural mechanisms that support defensive behaviors. For example, rats will freeze, or become motionless except for breathing, when presented with a stimulus, such as a tone, that had been paired with a painful unconditional stimulus, such as a foot shock. This conditional emotional response is thought to reflect an acquired fear of a previously neutral stimulus. Information about the conditional and unconditional stimuli converge in the amygdala, which has been shown to be important for the acquisition and expression of fear. Unlike in the case of eyeblink conditioning, in which a discrete motor response is learned, learned fear activates a number of different responses—freezing, as well as effects on heart rate, respiration, and digestion. The amygdala orchestrates this system through its outputs to brain stem structures. Like the cerebellum, the amygdala is not necessary for declarative learning, but rather these two structures are key elements of two different nondeclarative memory circuits.

**Molecular Mechanisms of Learning**

The hippocampus, cerebellum, and amygdala are critical structures supporting different types of learning as shown primarily through lesion studies. Each of these structures has been the subject of intense scrutiny in terms of the changes that are occurring at the level of the synapse to support learning. Many of these studies are done using in vitro techniques in which a slice of brain is removed and kept alive in an oxygenated bath. This preparation allows access to individual neurons and their inputs.

The most popular in vitro preparation in the past several years has been the hippocampal slice. Using this preparation, one can record neural activity in the different hippocampal regions, including the dentate gyrus. While responses can be recorded from single neurons, researchers often measure the population spike—the result of a population of action potentials in response to stimulation of the input to this region. By stimulating the perforant path in the slice, a clear population spike can be recorded in the dentate gyrus. An exciting development in cellular neuroscience was the finding of long-term potentiation (LTP) in the dentate gyrus. After an intense stimulation of the perforant path (called a tetanus), a small stimulus now elicits a much larger population spike than the same stimulus did before the tetanus. This potentiation of neural activity can last many hours, as long as the slice remains viable. LTP has been often touted as a neural model of memory—a strong or salient input gives rise to a long-lasting change in a structure that is known to be critical for memory processing. Although LTP is readily measured in the hippocampus, it is by no means confined to there, as long-term plasticity (either potentiation or depression of activity) has been measured in cortex, striatum, and cerebellum.

The tetanus results in LTP because the rapid frequency of pulses means that neurotransmitter continues to be released even when the postsynaptic neuron is already depolarized. This situation also occurs when a weak input is paired with a strong input that depolarizes the cell. A weak input is as effective as a tetanus in causing LTP if the postsynaptic cell is simultaneously depolarized. These conditions would occur naturally when a weak and strong stimulus occur together—whereas the weak stimulus alone cannot cause the neuron to fire, when it is paired with a strong stimulus it gains strength to the point in which it can subsequently cause the neuron to fire on its own. This idea, that temporal proximity of firing of a weak and a strong input leads to a change in synaptic strength of the weak input, was promoted by Donald Hebb in 1949 as a means for associative learning in the nervous system.

LTP has been a popular model system to study possible cellular and molecular bases for learning. The N-methyl-D-aspartate (NMDA) receptor that is abundant in the hippocampus appears to be critical for the activity-dependent plasticity that leads to LTP. This receptor is activated by the neurotransmitter glutamate, which is the primary excitatory neurotransmitter in the brain. When the neuron is at rest, this receptor is blocked by a positively charged magnesium ion, which prevents glutamate from activating the receptor. However, when the postsynaptic cell is depolarized, and thus becomes more positively charged, the magnesium ion moves away, and glutamate released from a cell conveying a weak input can now bind to this receptor. Activation of the NMDA receptor results in a cascade of events in the cell that do not occur after activation of other types of glutamate receptors. The results of NMDA-receptor activation include the synthesis of new proteins that change the structure of the synapse, making it more efficient, and the release of nitric oxide that travels back to the presynaptic neuron and facilitates the subsequent release of neurotransmitter. Both of these mechanisms may underlie learning in the intact
mammalian brain. Interfering with the NMDA receptor with drugs such as MK-801 severely disrupts memory formation, suggesting that there is a link between this receptor and memory function.

**Future Directions**

Progress in understanding the neural basis of memory has proceeded in parallel at several levels of analysis: the level of the anatomical structures, of neural circuits within those structures, of changes in neurons, and of the molecules involved. Although progress at each of these levels has been immense over the past few decades, there is much work to be done in bridging between the levels. For example, it is unclear whether the same principles that apply to neural changes for simple conditioned responses also apply to declarative learning. Whereas the site of plasticity for simple conditioned responses is usually localized, it appears that the storage of declarative memories is more distributed in the cortex. Also, while the mechanisms underlying long-term potentiation have been studied extensively in the laboratory, these may not be identical to changes that occur under physiological circumstances. Questions such as these are likely to feature prominently in behavioral neuroscience in the next decade.

*Barbara Knowlton*

**See also** Amnesia; Implicit Memory; Reinforcement Learning, Psychological Perspectives

**Further Readings**


**MEMORY AND KNOWLEDGE**

One important strand of psychological research on memory has been concerned with uncovering ways in which human memory can be unreliable, misleading, or even involve complete fabrication. In apparent contrast to this line of research, philosophical discussions of memory typically seek to give it a central, indispensable role in knowledge. This entry will review some of the accounts philosophers have given of the epistemology of memory, before briefly returning to the question as to the potential relevance of empirical research in psychology for such accounts.

**What Role(s) Does Memory Play in Knowledge?**

To bring out one way in which memory may be thought to be central to knowledge, consider the following argument attacking the so-called myth of the given. Its target is a position in epistemology known as classical foundationalism, according to which empirical knowledge must ultimately rest on a set of “basic beliefs” whose epistemic status does not, in turn, depend on that of other beliefs. Traditionally,
foundationalists’ favorite candidates for such basic beliefs have been beliefs about our own sensations. Suppose you have a certain visual experience on the basis of which you form the belief “There is a ripe tomato in front of me.” Arguably, the epistemic status of that belief turns on whether you are right to believe that things actually are as they visually appear to you. Yet, the foundationalist would claim that there is another belief you can form in this situation, whose epistemic status does not seem to turn on that of other beliefs, namely, the belief that you are, at any rate, having a red sensation.

Implicit in this line of thought is the idea that the mere having of a sensation, by itself, can put you into a position to have a belief about it—and this is what has been criticized as the “myth of the given.” Your having the belief “a red sensation is occurring” or even just “this type of sensation is occurring” seems to require that you can think of your current sensation as being of one type rather than some other, which, in turn, seems to require drawing on memory. If you can’t remember (and thus have beliefs about) any other sensations you could have instead, your putative beliefs about your current sensation will be devoid of content—there will be nothing in them that can distinguish that sensation from any other sensation. Thus, it looks as though not even beliefs about our own sensations can serve as “basic beliefs” in the foundationalist’s sense.

It is often said that, without memory, we would know very little, because any knowledge we might have through sensory experience would only last as long as the experience itself. The preceding example suggests that this statement might not go far enough in acknowledging the epistemic centrality of memory. Rather, if the kind of attack against the “myth of the given” sketched is along the right lines, memory plays a key role in our very ability to gain knowledge from experience. At the same time, however, there is of course also a sense in which memory, in turn, depends on a capacity for experience (or other capacities for acquiring knowledge). Memory is not itself a faculty for coming to know something; it is dependent on there being other such faculties. As it is sometimes put, memory is not a source of knowledge, or, if it is, it is a preservative, rather than a generative, source.

Memory and Justification

Many epistemological theories focus primarily on generative sources of knowledge, and, as a consequence, at least some of them have difficulties accounting for the distinctive epistemological significance of memory. For instance, there is an influential tradition in epistemology that is centered on the notion of justification, often associated with the thought that justification is a necessary condition for knowledge (as opposed to, say, mere true belief). Following this tradition, it is sometimes assumed that questions about the epistemological significance of memory are best approached by asking in virtue of what beliefs retrieved from memory (henceforth: memory beliefs) might count as beliefs the subject is justified in holding. On closer inspection, however, it is not obvious that this latter question best captures the role memory plays in knowledge.

Consider one possible answer to the question as to what justifies memory beliefs: that remembering that $p$ is itself a source of justification for the belief that $p$. In effect, the strategy behind this answer is to downplay the epistemological significance of the generative/preservative distinction: Memory may be preservative with respect to content—that is, it preserves beliefs acquired by some other means—but it is generative with respect to justification. This latter idea is typically spelled out in terms of the thought that there is a specific phenomenology attached to retrieving beliefs from memory (as opposed to, say, just guessing). There is an experience of recall, which can serve as a justification for believing that $p$.

One problem for this view is that it is by no means clear how exactly invoking the idea of a distinctive phenomenology of retrieval might help flesh out the idea that remembering is itself a source of justification. It is perhaps tempting to think that such experiences can play a similar role in the justification of memory beliefs as, say, visual experiences play in the justification of visually based beliefs. Yet, intuitively, the epistemic role of perceptual experiences has something to do with the fact that there is a sense in which those experiences directly present us with the very things our beliefs are about. In particular, it is the specific perceptual experience I have which makes it rational for me to form the specific beliefs I do. By contrast, the putative epistemic role of memory beliefs, on the view we have been considering, would have to be rather different. Insofar as there is a distinctive experience of recall, it seems to be the same experience that accompanies different instances of
memory retrieval, no matter what beliefs are being retrieved.

An alternative answer to the question of what justifies memory beliefs turns on the idea that memory is preservative not just with respect to the content of beliefs but also with respect to their justification. As normally understood, the suggestion here is not that, to be justified in believing that \( p \) (where the belief that \( p \) is a memory belief), I need to be able to remember the circumstances under which that belief was acquired and thus be in a position to rehearse my original justification for acquiring the belief. This would make all but a fraction of our beliefs epistemically problematic. Rather, as it is normally understood, the view at issue here is that beliefs held in memory in fact retain the justification with which they were originally acquired, even if the subject herself is no longer able to remember how she acquired the belief.

This view faces the problem that it is not obvious what exactly the idea of a belief’s retaining its justification comes to. This idea seems to presuppose that justification is something akin to a property of beliefs, that is, states (of believing that \( p \), or \( q \), etc.) that a subject is in over time. Yet, when the notion of justification is explicated in the epistemological literature, it is typically by means of examples in which a subject acquires a belief for the first time, or holds on to a belief in the presence of countervailing evidence. That is, the notion of justification is attached to aspects of the subject’s cognitive activity, something the subject does at a time.

As already indicated, perhaps the most basic worry about both of the views outlined previously is whether it is right to assume that the epistemic significance of memory is best framed in terms of the notion of justification. As against this assumption, some authors have held that it is actually more intuitive to think of the role that memory plays in knowledge in terms of the idea that memory frees the subject from the need to seek justification for certain of her beliefs. Any sort of sustained rational inquiry seems to presuppose that we can normally rely on beliefs we acquired earlier without constantly having to establish their epistemic credentials anew. Thus, it might be thought that there is a sense in which memory has a more fundamental epistemic role to play than can be explained by invoking the idea of memory generating or preserving justification for individual beliefs held in memory. Rather, on this view, the epistemic significance of memory needs to be seen within its wider role of making it possible for us to acquire extended bodies of knowledge.

“False Memories” and the Epistemology of Episodic Memory

The previously discussed considerations are all concerned (at least primarily) with factual or semantic memory, that is, the ability to retain knowledge of facts, concepts, or meanings that we learned about in the past, but not necessarily knowledge about the past itself. Yet, perhaps the first thing to come to mind when the issue of the relation between memory and knowledge is raised is the idea that memory plays a role specifically in our knowledge about the past. Philosophical discussion of this idea has centered primarily on the epistemology of event or episodic memory, that is, the type of memory for particular past events that we might express, for example, by saying “I remember seeing/doing x.” In particular, theorists have tried to reconcile two intuitions about this type of memory: that it plays a fundamental role in our knowledge of the past, and that it involves the having of memory images.

The idea of a memory image is meant to capture a sense in which recollecting specific events from one’s past life involves something akin to reexperiencing them. Yet, it has been argued that the having of a memory image—that is, a present occurrence—cannot constitute our most fundamental way of knowing about the past. The thought has been that if we did not have a more fundamental way of knowing about the past, not involving imagery, we would never come to connect present memory images with the past. This argument, though, seems to assume that the only role imagery might play in knowledge about the past is by serving as evidence on the basis of which we make judgments about the past. It can be avoided if we can make sense of an alternative way of viewing the epistemic role of memory images. Specifically, it has been suggested that, in the case of episodic memory, the subject’s ability to call up a memory image is itself the specific form her knowledge of the past takes.

It is in this context, in particular, that empirical work on the reliability of memory might be thought to raise challenges for the epistemology of memory. Space prohibits a proper review of the large variety of empirical studies in this area. However, very broadly, a general theme that emerges from much of
this research is that subjects’ memories about past events are susceptible to interference from information received some time after the relevant events took place. At the extreme, entirely false apparent memories can be “planted” in subjects by giving them misleading information. This clearly raises a general challenge for epistemologists to provide grounds for thinking that, by and large, we can nevertheless regard memory as reliable. More specifically, though, the intuition that episodic memory has a distinctive epistemological role to play seems to trade on the idea that there is an essential difference between simply remembering facts about a past event and recollecting the event itself, that is, having some more direct access to the past through having witnessed it. It is this idea, which is arguably a key ingredient of our commonsense understanding of memory, that might be seen to be under threat once it is clear just how much what we seem to remember having experienced can actually be the result of post-event construction.

Christoph Hoerl

See also Knowledge by Acquaintance; Memory, Interference With; Objects of Memory

Further Readings

Memory Recall, Dynamics

How do we search our memories to recall information that occurred in a given temporal context? In the laboratory, this basic question concerning human memory is addressed by asking people to study a sequence of individually presented items (typically words) and then to try to recall all of the items they can remember in any order. This task, first introduced by E. A. Kirkpatrick in 1894, is termed free recall.

By analyzing the order in which participants freely recall list items, one can gain considerable insight into the nature of the recall process. The analysis of recall dynamics in free recall reveals several striking regularities. This entry first reviews five major phenomena that govern the way people search their memories: the effects of recency, primacy, contiguity, forward asymmetry, and semantic proximity. Subsequent sections discuss how these phenomena occur both in the patterns of correct recalls and recall errors, as well as in the latencies measured between successively recalled items. This entry closes with a brief discussion of the theoretical implications of these phenomena.

Recency

In immediate free recall, participants are far more likely to begin recall with the final list item than with an item from any other list position (Figure 1A). This tendency persists for the first several responses, after which recalls tend to come from more distributed list positions. Participants’ tendency to begin recall at the end of the list has been strongly linked to the well-known recency effect—the increased probability of recalling items from the end of the list. The striking recency effect seen in the data (Figure 1B) is greatly reduced when participants are asked to perform an unrelated cognitive task, such as mental arithmetic, in between list presentation and the recall period (delayed recall). Although the recency effect is easily disrupted in delayed free recall, other manipulations that influence overall recall performance have little effect on the recency effect.

Dissociations between recency and pre-recency effects in recall have led some theorists to argue for a fundamental distinction between short-term and long-term memory. In this view, recency arises due to retrieval from a limited capacity short-term store (STS) whose contents are easily displaced by new information. In contrast, recall of pre-recency items arises from a search of associative memory, where associations between items reflect both newly formed associations between items that were together in STS
and long-standing associative knowledge concerning the items themselves. If, however, recency depends exclusively on the operation of STS, then one would not expect to find recency in continual distracter free recall—a task where participants perform a distracting task (e.g., mental arithmetic) after every list item, including the last one. According to the STS account of recency, the final distracter should greatly attenuate the recency effect, as in delayed free recall. However, in continual-distracter free recall, one observes a strong recency effect and participants are nearly as likely to initiate recall with the final list item as in immediate free recall. This “long-term recency” has been taken to support the view that recency reflects a more general forgetting process that operates at both short and long time scales.

**Primacy**

In addition to the recency effect, one also observes a *primacy effect* in free recall, whereby the first few list items are remembered better than items from the middle of the list. This is seen both in the overall probability of recalling list items and in an increased tendency to initiate recall with the first list item (Figure 1). The primacy effect is largely attenuated when participants are discouraged from rehearsing list items throughout list presentation. This is because early list items tend to be rehearsed more frequently than other items (they have more rehearsal opportunities), and they also tend to be rehearsed throughout the input sequence, thus giving them a recency advantage. Unlike the recency effect, the primacy effect is not reduced in delayed free recall.

**Contiguity**

Because in free recall the order of recall reflects the order in which items come to mind, recall transitions reveal the organization of memory for the list items. Consider, for example, that a participant has just recalled an item from serial position \( i \) and that the next recall is from serial position \( j \). To examine the effects of the temporal organization of the list on free recall transitions, one can measure the relation between recall probability and the lag between \( i \) and \( j \), defined as \( j - i \). This measure is called the conditional-response probability as a function of lag, or lag-CRP.

Given that the participant has just recalled the item from serial position \( i \), the lag-CRP indicates the probability that the next item recalled comes from serial position \( i + \text{lag} \) given the possibility of making a transition to that serial position. Lag-CRP analyses have shown that the *contiguity effect*, a tendency for participants to recall items from nearby in the list to the just-recalled item, and the *asymmetry effect*, a tendency for participants to recall items in the forward direction, are extremely robust properties of free recall.
Figure 2A illustrates these phenomena. Positive values of lag = (j − i) correspond to forward recalls; negative values of lag correspond to backward recalls. Large absolute values of lag correspond to words spaced widely in the list; small absolute values correspond to words spaced closely together in the list. The contiguity effect seen in these data also appears in the form of shorter interresponse times (IRTs) between recall of items from neighboring list positions. This can be seen in the conditional-response latency (lag-CRL) function shown in Figure 2B.

The contiguity effect in free recall is also related to participant’s overall ability to recall list items. For example, older adults, who recall significantly fewer correct items than do younger adults, exhibit significantly reduced contiguity effects. Moreover, the magnitude of each participant’s contiguity effect is positively correlated with that participant’s recall performance.

**Associative Asymmetry**

An interesting feature of the contiguity effect, as seen in Figure 2, is the strong forward asymmetry, with recall transitions being nearly twice as likely in the forward than in the backward direction. This tendency to make forward transitions is also seen in serial recall (where it is more pronounced) and in the pattern of errors observed in probed recall of serial lists. However, the forward asymmetry effect in free recall contrasts with the finding that recall of words studied in pairs (e.g., BOY–TREE, SHOE–CAR, etc.) is almost perfectly symmetrical, with participants exhibiting nearly identical rates of forward and backward recall (BOY retrieves TREE just as easily as TREE retrieves BOY), and with forward and backward recall being highly correlated at the level of individual pairs. It may be that temporally segregated word pairs (as in paired associate memory tasks) are more likely to be encoded as distinct experiences than neighboring words in a list. Associative symmetry may thus be a property of well-integrated pairs that is broken by interference among items from different list positions.

**Semantic Proximity**

Whereas the contiguity effect illustrates the temporal organization of memories, it is also well known that participants also make use of preexisting semantic associations among list items. This can be seen in people’s tendency to make recall transitions among semantically related items, even in random word lists that lack obvious semantic associates. This semantic proximity effect can be seen in Figure 3A, which shows how the probability of making a recall transition among two items increases with their semantic relatedness. Not only are people more likely to make recall transitions among semantically related items, but they also make those transitions more quickly than transitions among less strongly related items (Figure 3B). Both of these effects are
Memory Recall, Dynamics

**Recall Errors**

Temporal contiguity and semantic proximity not only dictate the dynamics of correct responses in free recall; they also influence the kinds of recall errors people make. When recalling a list of words, participants occasionally recall items that were not on the target list. By examining the dynamics of recall, one can show that these intrusion errors exhibit the same three basic properties described earlier. First, they tend to be items that were studied on recent lists. Second, they tend to be semantically related to the just recalled (correct) item. Third, when participants commit two intrusions from the same prior list, they tend to be items that were studied in neighboring list positions. This latter result is another manifestation of the contiguity effect. Thus, the same three principles that govern the dynamics of correct recalls also help to explain the kinds of recall errors that people commit.

**Source Clustering**

One can also show that people exhibit clustering as a function of encoding task. Sean Polyn and colleagues asked participants to make either size or animacy judgments on different list items. During free recall, participants not only exhibited temporal and semantic clustering effects; they also exhibited clustering of responses based on the task in which the words were studied. That is, following recall of an item that was given a size judgment at encoding, participants were more likely to recall another item that was given a size judgment. Furthermore, this task clustering effect interacted with temporal clustering, being greater for items presented at neighboring list positions.

One may wonder whether the entire recall process can be described as a sequence of probabilistic draws influenced by temporal contiguity and semantic clustering effects, or whether there are changes in the dynamics of recall process over the course of the retrieval period. In immediate free recall, the contiguity effect is larger and the semantic proximity effect is smaller for the first few responses and then increases/decreases to a stable state for subsequent recalls. In delayed free recall, however, the contiguity effect and semantic proximity effect are relatively stable throughout the recall period.

**Interresponse Times**

In 1970, Ben Murdock and Ron Okada showed that interresponse times between successive recalls

---

**Figure 3** Semantic proximity effect. (A) Conditional-response probability as a function of semantic relatedness. (B) Conditional-response latency as a function of semantic relatedness.

Note: Semantic relatedness is determined using latent semantic analysis (LSA), which derives relatedness from the co-occurrence statistics of words that appear in a large corpus of text. Error bars represent Loftus-Mason corrected error.
increase steadily throughout the recall period, growing as an exponential function of the number of items recalled. This increase in IRTs is highly predictive of recall termination—following an IRT of greater than 10 seconds, people rarely recall further items. The dynamics of recall also appear to be significantly affected by recall errors. Following intrusions or repetitions of already recalled items, people have a significantly increased tendency to either commit further errors or terminate recall, a pattern that is true at all stages of the recall process. Although the exponential increase in IRTs during the recall period has been argued to support a model of recall in which items are randomly sampled with replacement from a set of available responses, this account is falsified by the strong dependencies in sequences of responses, including the temporal and semantic clustering effects, as reviewed earlier.

Retrieved Context Theory
Whereas the contiguity effect can be easily accommodated within the view that neighboring items become associated when they co-occupy a short-term buffer (or working memory system) several studies are hard to reconcile with this classic information processing account. For example, Marc Howard and Michael Kahana found that separating items by an unrelated distractor task (mental arithmetic) did not disrupt the relative tendency to make transitions to neighboring items. This finding was further extended in 2008 by Marc Howard and colleagues who asked participants to free recall items from 48 previously studied word lists. Under these conditions, participants exhibited a significant contiguity effect even when making recall transitions among items that occurred on different lists. For instance, following recall of an item from list 5, participants were more likely to recall an item from lists 4 or 6 than from lists 3 or 7, and so forth. In 2008, Orin Davis and colleagues also found that in recalling lists of paired associates, recall errors exhibited a strong contiguity effect, extending across several intervening pairs. Because interpair rehearsal would be a major source of interference in this task, it is unlikely that the contiguity effect can be entirely explained on the basis of rehearsal strategies. Even in item recognition of lists of pictures, participants exhibit contiguity effects in which recognizing a picture makes it easier to subsequently recognize a picture studied in a nearby list position. On the basis of these results, the contiguity effect may be seen as reflecting a kind of mental time travel undertaken during memory search and retrieval. In recalling an item, the participant “travels back” to the time of its presentation, making it more likely that subsequent recalls will come from neighboring list positions. According to this view, contiguity arises due to a contextual retrieval process in which recalling an item reinstates its associated temporal contexts, which, in turn, serve as a cue for neighboring items.

Summary
By analyzing the dynamics of memory retrieval in free recall, one can see how the search of episodic memories is a highly cue-dependent process. Five major principles govern the way people recall lists of studied items. First, people tend to initiate recall with recently studied items. Subsequent responses continue to show a bias toward recent items, but this recency effect rapidly dissipates over the course of retrieval (Figure 1). Second, recall of a given item tends to be followed by recall of an item from a neighboring (contiguous) list position—a phenomenon known as the contiguity effect (Figure 2). Third, the contiguity effect exhibits a strong forward asymmetry effect, with forward transitions being approximately twice as common as backward transitions (Figure 2). Fourth, recall of a given item tends to be followed by recall of a semantically related item (Figure 3). These principles not only govern correct responses; they also govern the errors people make during recall. A fifth principle is the tendency to make transitions to early list items, as seen in the primacy effect. Because primacy is not always observed and largely reflects people’s use of rehearsal strategies, this principle may be considered secondary to the first four major phenomena described earlier. By studying the order of recall responses, and not just whether or not items are recalled, one can observe the striking effects of temporal contiguity and semantic similarity on the accuracy and timing of both correct recalls and recall errors. The study of recall dynamics thus allows us to characterize the basic associative processes operating in recall and to test theories of these associative mechanisms.

Michael Kahana and Jonathan Miller
Mental Action

One central concern in the philosophy of action is to provide an account of what it is that distinguishes the things that merely happen to people—events they undergo—from the things they genuinely do, where these latter events are actions of an agent. For example, there is a difference between an event of one’s arm moving that is due to an involuntary twitch, and the event of one intentionally raising one’s arm. When one intentionally raises one’s arm, the bodily movement involved is not simply a mere happening that one undergoes, but a bodily action one performs. It has been argued that the same sort of distinction can be marked in the mental domain: Some mental events are not simply mental occurrences that subjects undergo, but are, rather, mental actions they perform.

Some philosophers argue that acknowledging that the perspective one has on one’s mental life can be that of its agent may have significant implications for the epistemology, metaphysics, and phenomenology of mind. However, there is disagreement over the question of which aspects of our mental lives should be regarded as mental actions. For example, while some hold that our mental actions include judgings and decidings, as well as calculatings, reasonings, and tryings, others have argued that although there is such a thing as mental action, most of our thoughts, including our decisions, just happen. This entry will briefly explore these issues.

The Scope of Mental Action

Thinking about something involves the occurrence of mental acts that are individuated by their propositional contents. For example, when one makes a conscious judgment, one judges that such and such is the case. What one judges to be the case can be of its agent may have significant implications for the epistemology, metaphysics, and phenomenology of mind. However, there is disagreement over the question of which aspects of our mental lives should be regarded as mental actions. For example, while some hold that our mental actions include judgings and decidings, as well as calculatings, reasonings, and tryings, others have argued that although there is such a thing as mental action, most of our thoughts, including our decisions, just happen. This entry will briefly explore these issues.

The Scope of Mental Action

Thinking about something involves the occurrence of mental acts that are individuated by their propositional contents. For example, when one makes a conscious judgment, one judges that such and such is the case. What one judges to be the case can be of its agent may have significant implications for the epistemology, metaphysics, and phenomenology of mind. However, there is disagreement over the question of which aspects of our mental lives should be regarded as mental actions. For example, while some hold that our mental actions include judgings and decidings, as well as calculatings, reasonings, and tryings, others have argued that although there is such a thing as mental action, most of our thoughts, including our decisions, just happen. This entry will briefly explore these issues.

The Scope of Mental Action

Thinking about something involves the occurrence of mental acts that are individuated by their propositional contents. For example, when one makes a conscious judgment, one judges that such and such is the case. What one judges to be the case can be of its agent may have significant implications for the epistemology, metaphysics, and phenomenology of mind. However, there is disagreement over the question of which aspects of our mental lives should be regarded as mental actions. For example, while some hold that our mental actions include judgings and decidings, as well as calculatings, reasonings, and tryings, others have argued that although there is such a thing as mental action, most of our thoughts, including our decisions, just happen. This entry will briefly explore these issues.

The Scope of Mental Action

Thinking about something involves the occurrence of mental acts that are individuated by their propositional contents. For example, when one makes a conscious judgment, one judges that such and such is the case. What one judges to be the case can be of its agent may have significant implications for the epistemology, metaphysics, and phenomenology of mind. However, there is disagreement over the question of which aspects of our mental lives should be regarded as mental actions. For example, while some hold that our mental actions include judgings and decidings, as well as calculatings, reasonings, and tryings, others have argued that although there is such a thing as mental action, most of our thoughts, including our decisions, just happen. This entry will briefly explore these issues.
formed the intention to $F$, one might choose to express that intention, but the mental acts of judging that $p$ and deciding to $F$ cannot themselves be intended.

One response to this line of thought is to claim that when one is engaged in directed thinking, such as trying to figure out an arithmetical problem, although one does not intend to think a thought with the content that $p$, intention may still have a significant role to play in one's mental activity, for one's intention may be to think a thought that stands in a certain relation to other thoughts or contents. Although one does not decide to judge that $p$, and one does not decide to decide to $F$, one may be able to decide to determine (or attempt to determine) whether $p$, and one may be able to decide to decide (or attempt to decide) whether to $F$. The conclusion that some draw from this is that although mental acts like judging and deciding are not themselves mental actions, mental action may sometimes play an important role in explaining their occurrence.

An alternative response is to reject the assumption that a mental event can only be a mental action if the content of the mental event is intended. For example, according to Christopher Peacocke, for a mental event to be a mental action, it must consist of an event that either is, or constitutively involves, a trying, and Peacocke has argued that tryings should be distinguished from prior intentions. Peacocke argues that conscious events of deciding and judging can be caused by such events of trying, and so can be mental actions. He has suggested that such mental actions have the phenomenology of doing something, rather than involving the phenomenology of being presented with something as being the case, as in perception, or as something occurring to one, as in unintended imagination.

**Mental Action and Self-Knowledge**

For Peacocke, knowing what one is consciously thinking will often involve knowing what mental actions one is performing. Peacocke's account of how we standardly know our own actions appeals to the occurrence of events of action-awareness that are not beliefs and that have a first-person, present tensed content of the form “I am doing such-and-such now”—for example, “I am judging that $p$ now.” This action-awareness is standardly brought about by an event of trying that causes the action that the action-awareness represents. According to Peacocke, the distinctive way in which a subject comes to know of his own actions is by taking such an apparent action-awareness at face value. He argues that this distinctive action-awareness exists for mental actions, as well as for bodily actions.

Other accounts of how one knows what one is doing when performing an action appeal to the idea that our intentions can embody a form of self-knowledge—practical knowledge of our intended actions. According to this view, intended action is behavior that realizes the agent's knowledge of it. The suggestion is that this can accommodate the intuition that one does not normally need to find out by observation or inference what one is doing when performing an intentional action, for performing the action was one's own idea to begin with. On this view, one's knowledge of what one is thinking may sometimes be explained, in part, by the distinctive epistemic role of one's intentions in thinking.

**Mental Action and Wakeful Consciousness**

Brian O'Shaughnessy appeals to the idea that there is a distinctive form of self-knowledge that accompanies one's mental actions in arguing for the claim that mental action has a crucial role to play in an account of the state of wakeful consciousness in the self-conscious. He argues that the progression of the stream of conscious thought and imagination is, in the awake self-conscious subject, distinctive. The respect in which it is distinctive is connected with the variety of self-knowledge that accompanies it, and the relevant form of self-knowledge is linked with the idea that the “mental will” is operative. According to O'Shaughnessy, the awake self-conscious subject (as opposed to one who is dreaming) is able to make sense of what is happening in a certain domain of his mental life insofar as he is able to make sense of what he is up to, and the variety of self-knowledge involved is importantly linked with the idea that the perspective he has on this aspect of his mental life is that of its agent.

Matthew Soteriou

See also Action and Bodily Movement; Philosophy of Action; Self-Knowledge

**Further Readings**

Mental Causation

Mental causation is the causation of physical effects by mental causes. The paradigm case of mental causation is the causation of someone’s bodily movement by a mental state or event of hers. The belief that mental causation exists is deeply rooted in common sense. It seems uncontroversial to say, for instance, that a sudden pain caused Jones to wince, or that Smith’s thirst caused him to have a drink. Nevertheless, explaining how the mind can have physical effects has proven a challenge for philosophers of mind. For physical effects already have physical causes, which threatens the claim that they also have mental causes. The problem is most pressing for positions according to which the mind is not itself physical. However, recent decades have also seen a debate over whether the view that the mind is physical can adequately explain mental causation.

History

The existence of mental causation was generally considered uncontroversial by ancient philosophers. For instance, both Plato and Aristotle, although differing in their views about the nature of the mind, held that agents’ mental states need to be invoked in order to give causal explanations of some of their bodily movements.

The modern debate about mental causation can be traced back to René Descartes and the controversy about his theory of the mind. Descartes held that minds and bodies are two radically different kinds of substance: Minds are substances that are thinking and not spatially extended, whereas bodies are substances that are spatially extended and not thinking. (By a substance in general, Descartes understood something that exists and whose existence does not depend on anything else.) In correspondence, Princess Elizabeth of Bohemia complained to Descartes that she found it unintelligible how his theory could allow minds to cause the motion of bodies. She held that bodies could only be moved by things in spatial contact with them, which ruled out minds as causes of bodily movements because they lacked the required spatial attributes. Although he never resolved the dispute with Princess Elizabeth, Descartes later developed a theory that identified the pineal gland as the locus of mind-body interaction. By moving the pineal gland, he claimed, the mind affects the motion of our animal spirits (an air-like kind of matter), which communicate the impulse to our muscles via the nerves.

Gottfried Wilhelm Leibniz criticized Descartes’ position for being at odds with physics. He held that the law of conservation of momentum was violated if minds affected the motion of bodies in the way envisaged by Descartes. Leibniz’s own position denied mind-body interaction altogether. According to his view, different substances never interact, but God created them so that their histories unfold independently in perfect, preestablished harmony.

The Argument From the Causal Completeness of the Physical

In the 21st century, virtually no one endorses Leibniz’s doctrine of preestablished harmony or Princess Elizabeth’s conception of the motion of bodies. Still, most contemporary philosophers share the spirit of their objections to Descartes, which demands that mental causation fit into our picture of the physical world. One element of this picture is the principle of the causal completeness of the physical, which says that every physical effect has a physical cause (this principle is also called “causal closure of the physical”):

Completeness: Every physical event that has a cause has a physical cause.

Completeness is the starting point for an influential argument about mental causation. It seems that if an event has a physical cause, this cause is sufficient to bring the event about, which rules out that any nonphysical causes are involved. This idea is expressed by the following principle:

Non-redundancy: If an event has a physical cause, it does not have any nonphysical causes.

We may add to our assumptions the commonsensical view that some mental events, such as Jones’s pain or Smith’s thirst, have physical effects:
Mental causes: Some mental events have physical effects.

From completeness, non-redundancy, and mental causes, it follows that some mental events, namely those that have physical effects, are physical causes and thus are themselves physical events. Because it is implausible that there should be a difference in kind between those mental events that have physical effects and those that do not, the conclusion generalizes to the claim that all mental events are physical events.

Objections to the Argument

If one denies the conclusion of the argument from completeness, non-redundancy, and mental causes, one has to reject at least one of its premises. If one rejects mental causes, one has to hold that no mental events have physical effects; this view is called epiphenomenalism. Accepting epiphenomenalism comes at a price, as it requires abandoning the intuitively plausible claim that some of our mental events cause bodily movements. This has far-reaching consequences: Given that performing intentional actions requires that intentions and desires cause bodily movements, it follows from epiphenomenalism that we never perform intentional actions.

Whereas completeness seems to have a good standing, some philosophers have taken issue with the assumption of non-redundancy. Defenders of non-redundancy typically reply that giving up non-redundancy means accepting that some physical events are overdetermined by physical and non-physical events; it is implausible, they hold, that overdetermination is so widespread a phenomenon as accounting for all cases of mental causation would require. Whether such widespread overdetermination would be objectionable is a matter of controversy. The issue is complicated by the fact that overdetermination can be read in two ways. It can either simply mean that an effect has two causes, or it can refer to a case with a specific causal structure analogous to the case of a firing squad, where two shots are individually sufficient to bring about the victim’s death.

Physicalism

If one accepts the generalized conclusion of the argument from completeness, non-redundancy, and mental causes, one has to accept the claim that all mental events are physical events. This claim can be spelled out in different ways.

Type Identity

According to the type identity theory, every type of mental event is identical to a type of physical event. For instance, a proponent of the type identity theory might hold that pain is identical to a certain type of neural event. The type identity theory has been criticized for reasons independent of mental causation. It seems that mental events are multiply realizable. For instance, for some animals, pain may coincide with a physical event that is of a different type from the neural event occurring in humans when they are in pain; there might even be possible beings that can be in pain while their physiology differs radically from that of any animals we know. Multiple realizability contradicts the type identity theory. For if pain is identical to a certain type of neural event, pain and this neural event are one and the same type of event, so that, necessarily, if a pain event occurs, so does an event of the neural type.

Token Identity

Instead of accepting the type identity theory, one may opt for the weaker theory that identifies each token mental event, that is, each particular occurrence of a mental event, with a given token physical event. This so-called token identity theory allows for multiple realizability, since the different tokens of pain events, say, may be identical to token physical events of different types.

The most influential token identity theory has been Donald Davidson’s theory of anomalous monism. While identifying token mental events with token physical events (hence “monism”), Davidson denies that there are strict laws relating the mental and the physical (hence “anomalous”). He holds that causation requires strict laws, and that token mental events can be causes or effects because they fall under physical descriptions and hence are subject to the strict laws of physics. It has been objected against anomalous monism that, while it allows token mental events to be causes, it does not explain how token mental events can be causally efficacious by virtue of their mental properties. It has been argued that because it is only mental events’ physical properties that matter for their causal relations, according to anomalous monism, their mental properties, such
as the property of being a pain event, are rendered causally irrelevant.

**Supervenience**

Currently the most common view according to which mental events are physical events is *supervenience physicalism*. In a standard formulation, this view says that whenever a mental event of a certain type occurs, this is by virtue of the occurrence of some type of physical event that necessitates the occurrence of the mental event. Like the token identity theory, supervenience physicalism is compatible with the multiple realizability of mental events. For it is consistent with supervenience physicalism that different occurrences of a certain type of mental event are due to occurrences of different types of physical events, provided that an event of the mental type could not have failed to occur given the occurrence of an event of any of these physical types.

With respect to mental causation, supervenience physicalism faces the so-called exclusion problem. Suppose that a certain mental event occurs. Given supervenience physicalism, this occurrence is due to, and necessitated by, the occurrence of some physical event. It seems that any putative physical effects of the mental event will already be caused by the physical event. This, however, calls the causal efficacy of the mental event into question. The rationale behind the final step of this argument resembles the non-redundancy principle: If a physical effect already has a physical cause, other events simultaneous with this cause seem to be rendered causally irrelevant. Some philosophers accept that the exclusion problem makes mental events causally irrelevant if supervenience physicalism is true. Others contend that, unlike in the case of physical versus nonphysical events, supervenient mental events do not compete with their underlying physical events for causal efficacy. The latter philosophers are often motivated by a desire to prevent generalized epiphenomenalism. For if the exclusion problem is genuine, it generalizes beyond mental events and threatens the causal efficacy of other entities that plausibly supervene on the physical, such as the events and properties described in chemistry and biology.

**Content Externalism**

A further problem of mental causation arises for the widely held position of content externalism, according to which some mental states and events have contents that depend not merely on what is going on in the subject’s head but also on the subject’s environment. For instance, a content externalist might hold that the content of Smith’s belief that he is holding a glass of water is partly due to the fact that Smith has been in causal contact with water in his environment. It seems desirable to be able to say that contentful mental states and events can have physical effects and that their contents play a role in their causal efficacy. However, on the face of it, it seems that the causes of an agent’s bodily movements are internal to the agent and independent of the environmental factors relevant for the content of the agent’s thoughts. How content externalism might be reconciled with the causal relevance of mental content is a matter of ongoing controversy among philosophers of mind.

*Thomas Kroedel*

**Further Readings**


Mental Effort

Mental effort refers to the intensity of a person’s engagement with a cognitive task. High-effort situations are associated with high load on attention and executive control mechanisms, as distinct from situations in which cognitive processing is more automatic. Committing more effort can improve task performance. Applying effort to a focal task may also detract from performance on secondary tasks. Effort can be associated with a subjective sense of difficulty, strain, or work. Effort also tends to be accompanied by bodily arousal and is sometimes measured via increases in heart rate, pupil diameter, blood pressure, or galvanic skin response.

For most of its history, mental effort has remained an elusive and slippery construct, but the application of modern neuroscientific methods coupled with the development of formal theory promises to lend it greater rigor.

This entry describes the role of mental effort in theories of cognitive function, addresses how people strategically decide whether and how to devote effort to tasks, discusses how people use their experience of mental effort to make other inferences, and reviews relevant findings in neuroscience.

Mental Effort and Its Place in Theories of Cognition

Cognitive Resources and Capacity Mobilization

Traditional, resource-based models regard cognitive performance as dependent on a limited supply of cognitive resources, energy, or channel capacity. Within this framework, devoting more effort to a task involves allocating more general-purpose resources to it. For tasks that are more difficult, successful performance draws more resources away from other ongoing activities. Thus, the amount of effort devoted to a focal task can be measured by the performance decrement on secondary tasks.

In an early, influential treatment of the subject, Daniel Kahneman suggested that mental effort might involve not only resource reallocation but also modulation of the attentional system’s total capacity. On this view, mental effort—and the associated physiological arousal—reflects the transient expansion of an elastic pool of available resources. Viewing effort as capacity recruitment can explain why markers of physiological arousal respond both to cognitive demands and to incentives.

Effects of Effort on Performance

Increased effort can improve performance on a variety of target-detection, memory, problem solving, and decision tasks. This can be demonstrated by showing that performance varies together with task-related incentives. For instance, individuals dividing their attention between simultaneous tasks tend to perform better on the task with the greater incentive.

Some tasks are relatively insensitive to changes in effort (e.g., easy tasks can often be performed well with little commitment of effort). In other cases, effort can even be detrimental. If a task involves well-practiced, automatic physical skills (e.g., golf putting), arousal and focused attention can cause “choking.” Effort can impair types of problem solving that depend on associative processing or sudden insight. Additional effort can also be harmful if physiological arousal is already high (e.g., in a test-taking environment), a principle known as the Yerkes-Dodson law.

Self-Regulatory Depletion

Although resource-based frameworks have traditionally emphasized competition among concurrent activities, exertion of mental effort can also impair later performance. Roy Baumeister and colleagues have proposed that the mental resources underlying self-regulation not only are limited but also can be temporarily depleted. For example, if someone initially performs a task in which he or she must override impulses, that person might show reduced self-control or perseverance in a subsequent activity. Such findings have been used to support a “muscle” metaphor, wherein the ability to exert effortful self-control is subject to short-run fatigue.

Controlled Versus Automatic Processing

Mental effort is associated with processes that require executive control, as distinct from more automatic processes. For example, it is effort demanding to maintain information in working memory in the face of interference, or to override automatic responses. Controlled processing is typically assumed to proceed slowly, serially (rather than in parallel), and to facilitate pursuit of goals that are
poorly supported by the immediate environment. Tasks that are controlled and effort demanding when they are initially learned, such as searching for a target in a field of distracters, can come to demand less effort after extensive practice.

Research on judgment and decision making similarly recognizes a distinction between two broad categories of decision processes that closely parallels the distinction between automatic and controlled processing: “System 1” processes are relatively fast and automatic, whereas “system 2” processes are slower, more deliberative, and associated with a greater sense of effort.

Effort Costs and Effort Allocation

When physical exertion is involved, people generally choose the least demanding paths to their goals. There is evidence that a similar principle applies to mental effort. For a given activity, people tend to weigh the cost of effort against its expected payoff.

Effort seems to be recruited on an as-needed basis. In psychophysiology experiments, arousal tends to increase as people prepare for moderately difficult tasks (where effort will likely help), but not when they prepare for tasks that are trivially easy or impossibly difficult (where effort is unlikely to change performance substantially). These types of findings suggest that effort allocation involves strategic decisions.

Models of attention and executive control often posit an internal mechanism for monitoring ongoing levels of cognitive demand. This monitoring mechanism regulates the engagement of effort, attention, and control; engagement of these functions is high when needed, and low otherwise. Effortful cognitive control is treated in such models as if it carries a cost, not being engaged unless it is expected to bring gains.

The Nature of Effort Costs

One possible explanation for effort costs is metabolic consumption. If mental effort consumes energy, minimizing effort might help conserve nutritive resources. However, this is a matter of debate; effortful cognitive processing seems to bring about relatively small changes, if any, in the brain’s total glucose consumption. Alternatively, effort might be viewed as involving an opportunity cost. If there is a limited capacity for cognitive control, then capacity devoted to one task cannot be devoted to another. Yet another possibility is interference: Trying to carry out many tasks at once can cause cross-talk in processing and responding. Effort may be the brain’s way of limiting the potential for such interference.

Effort and the Selection of Decision Strategies

Tradeoffs between effort and accuracy have been discussed extensively in the field of decision making. When someone making a decision faces many sources of information and possible responses, there are many ways they could integrate the information and arrive at a decision. Simpler strategies might save effort, while reducing the likelihood of making the best possible choice.

There is evidence that decision makers often rely on simplifying heuristics, such as considering only a subset of the relevant information or the available alternatives. Several researchers, including John Payne and colleagues, have proposed that people choose strategies adaptively, rationally evaluating the expected accuracy and effort of each possible strategy. A simplifying decision strategy (e.g., focusing on only a single source of information) might be chosen if it minimizes the costs of effort with only a modest loss of accuracy.

Personality

Individuals differ in their inclination to engage in mentally effortful activities or to process information deeply; for example, people vary on a scale of need for cognition. Individuals also vary in measures of cognitive ability, such as general fluid intelligence and working memory span. Both motivation and ability are likely to influence the costs and benefits attributed to mental effort.

Metacognition: Interpreting Effort

How people interpret and evaluate their own mental states is known as metacognition. One form of metacognition is the use of the experience of mental effort as a basis for predictions and factual inferences. This has been studied experimentally by manipulating disfluency, the subjective sensation that information requires effort to perceive, remember, or process. Subtle manipulations of disfluency (such as degrading text legibility) can influence a variety of judgments. Disfluency can cause people to judge that products are less valuable, cities are less
populous, pictures are less likeable, and that aphorisms are less likely to be true.

In a notable example, Norbert Schwarz and colleagues asked participants to evaluate their own level of assertiveness. Before making this judgment, some participants were asked to recall 12 examples of their own assertive behavior; other participants were asked for 6 examples. Members of the 12-example group then described themselves as less assertive, despite having recalled a greater amount of supportive information. These participants presumably inferred, from the sense of mental effort involved with generating 12 examples, that such examples were scarce.

Experiences of mental effort can also inform judgments of learning. Information might be judged as poorly learned if it takes effort to interpret and recall; as a result, a person might devote additional study time to the material. Conversely, if information feels easy to interpret, a person might conclude (perhaps inaccurately) that it does not need to be studied further.

Mental Effort and the Central Nervous System

As noted earlier, mental effort is often measured using physiological indices of autonomic nervous system activity. Mental effort is also associated with specific patterns of brain activity.

In neuroimaging experiments, tasks that require controlled processing consistently activate a network of brain regions, including lateral prefrontal, anterior cingulate, and parietal cortices. This collection of regions has been referred to variously as the cognitive control, executive attention, or task-positive network. The network responds to numerous manipulations of difficulty, including working memory demands, task novelty, and response override.

A separate network of brain regions, including ventromedial prefrontal and posterior cingulate cortices, often shows activity reductions during tasks that demand effort. This system is sometimes called the default mode or task-negative network. An active hypothesis is that these regions support episodic or self-reflective thoughts, and that such thoughts decline in frequency when individuals commit mental effort to a task.

Mental effort is also associated with specific neuromodulators, including dopamine and norepinephrine. Both of these neurotransmitters are released by brain stem nuclei with diffuse cortical projections, exerting broad influence on cortical dynamics. Dopaminergic modulation influences the executive attention network; dopamine levels relate to incentives, response vigor, physical effort exertion, and control-related cognitive processes such as working memory. Norepinephrine levels influence focused attention and are related to indices of physiological arousal such as pupil diameter.

A neuropsychological case study provides striking evidence linking the cognitive control network to physiological and subjective aspects of mental effort. The patient, reported by Lionel Naccache and colleagues, had a large medial frontal cortical lesion that included anterior cingulate cortex. She could successfully perform demanding cognitive tasks, but these tasks evoked neither their usual physiological response nor any subjective sense of effort. This suggests that the subjective and physiological characteristics of effort can be dissociated from response generation processes.

Joseph T. McGuire,
Jonathan D. Cohen, and
Matthew M. Botvinick

See also Attention, Resource Models; Automaticity; Metacognition and Education; Two System Models of Reasoning; Working Memory

Further Readings

META-ANALYSIS

Meta-analysis is the quantitative cumulation of scientific evidence. There has long been pessimism in the younger social, behavioral, educational, and biomedical sciences in that our progress has been slower and less orderly than we would like, at least when compared to the progress of older, more programmatic sciences, such as physics and chemistry. In other words, the more recent work in physics and chemistry seems to build directly on the older work of those sciences, whereas the more recent work of the social, behavioral, educational, and biomedical sciences seems often to be starting from scratch. Those who have looked closely at the issue of cumulation in the physical sciences have pointed out that these disciplines have ample problems of their own. Nonetheless, in the matter of cumulating evidence, the social, behavioral, educational, and biomedical sciences have much to be modest about.

Limited success in the process of cumulation does not seem to be due to lack of replication or to the failure to recognize the need for replication. Indeed, there are many areas of the social, behavioral, educational, and biomedical sciences for which the results of many studies, all addressing essentially the same question, are available. Our summaries of the results of these sets of studies, however, have not been nearly as informative as they might have been, either with respect to summarized significance levels or, more important, with respect to summarized effect sizes, that is, the magnitudes of the effects examined. Even the best reviews of research by the most sophisticated scholars have been primarily qualitative narratives and have rarely told us much more about each study in a set of studies than the direction of the relationship between the variables investigated, and often not even that, and whether or not a given significance level was attained.

This state of affairs is beginning to change, however. More and more reviews of the literature are moving from the traditional literary approach to quantitative approaches to research synthesis described in an increasing number of textbooks of meta-analysis. The goals of these quantitative approaches of meta-analysis are to help us discover what we have learned from the results of the studies conducted and to help us discover what we have not yet learned.

In what follows, this entry defines the concept of “research results,” briefly examines the history of meta-analysis, defines the concept of “successful replication,” and concludes by comparing a more traditional view of replication success with a newer, probably more useful view.

Defining Research Results

Before we can consider various issues and procedures in the quantitative cumulation of research results, we must become quite explicit about the meaning of the concept “results of a study.” It is easiest to begin with what we do not mean. We do not mean the prose conclusion drawn by the investigator and reported in the abstract, the results, or the discussion section of the research report. We also do not mean the results of an omnibus $F$ test with $df > 1$ in the numerator or an omnibus $\chi^2$ test with $df > 1$, that is, tests of statistical significance that leave unspecified exactly what differences were found.

What we do mean is the answer to the question: What is the relationship between any variable $X$ and any variable $Y$? The variables $X$ and $Y$ are chosen with only the constraint that their relationship be of interest to us. The answer to this question should normally come in two parts: (a) the estimate of the magnitude of the relationship (the effect size), and (b) an indication of the accuracy or reliability of the estimated effect size (e.g., as indexed by a confidence interval placed around the effect size estimate). An
alternative, or better, an addendum, to the second part of the answer is one not intrinsically more useful, but one more consistent with the existing practices of researchers: the significance level of the difference between the obtained effect size and the effect size expected under the null hypothesis (usually an effect size of zero).

Because a complete reporting of the results of a study requires the report of both the effect size and level of statistical significance, it is useful to make explicit the relationship between these quantities. The general relationship is given by the following:

\[
\text{Significance Test} = \text{Effect Size} \times \text{Study Size}
\]

In other words, the larger the study is, in terms of the number of sampling units being studied, the more significant the results will be. This is true unless the size of the effect is truly zero, in which case a larger study will not produce a result that is any more significant than a smaller study. Effect sizes of exactly zero, however, are rarely encountered.

**Meta-Analysis: A Historical Note**

We are inclined to think of meta-analysis as a recent development, but it is older than the famous \( t \) test, which dates back a hundred years. We can simultaneously describe the early history of meta-analysis, while providing a classic illustration of the meta-analytic enterprise. In 1904, Karl Pearson collected correlation coefficients (called \( r \)); there were six of them, with values of .58, .58, .60, .63, .66, and .77. The weighted (by sample size) mean of these six correlation coefficients was .64, the unweighted mean was .63, and the median was .61. Pearson was collecting correlation coefficients because he wanted to know the degree to which inoculation against smallpox saved lives. His own rough and ready summary of his meta-analysis of six studies was that there was a .6 correlation between inoculation and survival—a truly huge effect. In practical terms, a correlation of that magnitude describes a situation in which inoculation increases the survival rate from 20% to 80%.

When Karl Pearson quantitatively summarized six studies of the effects of smallpox inoculation, a meta-analysis was an unusual thing to do. Recently, however, there has been an explosion of meta-analytic research syntheses such that a rapidly increasing proportion of all reviews of the literature are in the form of quantitative reviews, that is, meta-analyses. Despite its increasing frequency in the literature, however, meta-analysis is not without controversy and criticism (see Further Readings).

To gain a deeper understanding of meta-analytic procedures, it will be useful to consider the concept of replication. Meta-analysis, after all, involves summarizing or synthesizing studies that are broadly thought of as replications. It is important to note that studies typically included in meta-analyses are not replications in a narrow sense. Rather, they examine the same underlying relationships even if their independent and dependent variables are operationally defined in different ways. For example, in a meta-analysis of the effects of psychotherapy, the independent variables might be behavior therapy versus placebo, or psychodynamic behavior therapy versus placebo, or cognitive behavior therapy versus placebo. The dependent variables might be patients’ scores on a standardized paper and pencil measure of psychological health, or the ratings by observers of the effectiveness of patients’ interpersonal interactions, or the patients’ report of the degree to which they are experiencing an improvement in their enjoyment of life.

**Defining Successful Replication**

There is a long tradition in psychology of urging replication of each other’s research. Although we have been very good at calling for such replications, we have not been very good at deciding when a replication has been successful. The issue we now address is this: When shall a study be deemed successfully replicated? Ordinarily, this is taken to mean that in a new study at time 2, a null hypothesis that has been rejected at time 1 (i.e., was found significant) is rejected again, and with the same direction of outcome. When one study is significant and the other is not, we have a “failure to replicate,” but such “failures” may be quite misleading. Let us consider an example.

**Pseudo-Failures to Replicate**

**The Saga of Smith and Jones**

Smith has published the results of an experiment in which a certain treatment procedure was predicted to improve performance. She reported results significant at \( p < .05 \) in the predicted direction. Jones published a rebuttal to Smith, claiming a failure to
replicate. Both had an effect size $r$ of .24. But Smith had 80 participants and Jones had only 20. In this type of situation, it is often the case that, although the $p$ value associated with Smith’s results is smaller than Jones’s (i.e., more significant), the studies were in quite good agreement as to their estimated sizes of effect as defined by $r$; the correlation (technically, the point biserical correlation), between group membership (coded 0 or 1) and performance score (a more continuous score). Thus, studies labeled as “failure to replicate” may turn out to provide quite strong evidence for the replicability of the claimed effect.

**On the Odds Against Replicating Significant Results**

A related error often found in the behavioral, educational, biological, and social sciences is the implicit assumption that if an effect is “real,” we should expect it to be found significant again on replication. Nothing could be further from the truth.

Suppose there is, in nature, a real effect of treatment with a true magnitude of $r = .24$, equivalent to a difference in success rates of 62% versus 38%. Further suppose an investigator studies this effect with an $N$ of 64 participants or so, giving the researcher a level of statistical power of .50, a very common level of power for behavioral researchers of the past 45 years. Even though an $r$ of .24 can reflect a very important effect, there is only one chance in four ($p = .25$) that both the original investigator and a replicator will get results significant at the .05 level; that is, the probability (power) for the first study ($p = .50$) is multiplied by the probability for the second study ($p = .50$) to yield $.50 \times .50 = .25$. If there were two replications of the original study, there would be only one chance in eight ($p = .125$) that all three studies would be significant (i.e., $p = .5 \times .5 \times .5 = .125$), even though we know the effect in nature is very real and very important.

**Contrasting Views of Replication**

The traditional, less useful view of replication success has two primary characteristics: (a) It focuses on significance level as the relevant summary statistic of a study, and (b) it makes its evaluation of whether replication has been successful in a dichotomous fashion. For example, replications are successful if both or neither $p < .05$, and they are unsuccessful if one $p < .05$ and the other $p > .05$. Behavioral researchers’ reliance on a dichotomous decision procedure has been well documented. In this dichotomous procedure, differences between $p$ levels are all regarded as trivial except the difference between a $p \leq .05$ and a $p > .05$, or some other critical level of significance at which we have decided to “reject the null hypothesis.” This dichotomous approach to significance testing has been increasingly criticized, for example, by the American Psychological Association’s Task Force on Statistical Inference.

The newer, more useful view of replication success has two primary characteristics: (a) a focus on effect size as the more important summary statistic of a study, with a relatively more minor interest in the statistical significance level, and (b) an evaluation of whether replication has been successful made in a continuous, not dichotomous, fashion. For example, two studies are not said to be successful or unsuccessful replicates of each other, but rather the degree of failure to replicate is indexed by the magnitude of difference between the effect sizes (e.g., $r_s$) obtained in the two studies.

Robert Rosenthal

**Further Readings**


Lying in bed, Kendall is doing science homework. Kendall usually enjoys science and believes she can succeed on most of her classwork. She is particularly motivated to learn the planets, because her fifth-grade teacher has just discussed the solar system. Kendall repeats every planet name while looking at its picture, and then she decides to cover up each name to test herself. She remembers most of them but forgets Saturn and Venus, so she decides to spend more time memorizing their names.

In this scenario, Kendall is relying on cognitive and metacognitive processes to complete her homework. Cognition refers to mental activities and mental representations, such as Kendall’s underlying memories for the planet names. Metacognition refers to people’s thoughts (or cognitions) about their cognitions, such as Kendall’s understanding that testing herself may help her identify planets that have not been stored in memory. Accordingly, her understanding of cognition (metacognition) may lead her to use effective strategies and, in turn, improve her memory (cognition). This entry discusses the role of metacognition in education.

Components of Metacognition

Metacognition is not a unitary construct; it has three general components: knowledge about cognition, monitoring of cognition, and control of cognition. Metacognitive knowledge includes any knowledge or beliefs—whether they are valid or invalid—about how our minds operate. An expert on cognition would have detailed metacognitive knowledge about how the mind operates, whereas many students may have rudimentary knowledge about how the mind operates. For instance, many students do not know how to study most effectively or how to optimally solve problems. Monitoring refers to assessing the current state of any cognitive process, such as when Kendall attempted to monitor her progress by testing her memory for each planet. In doing so, she was able to use the outcome of the test—for example, whether she recalled the correct name and how quickly she recalled it—to infer how well she had learned the planet’s name. Control refers to regulating any aspect of cognition, such as beginning a new activity, doing the same one in a different manner, or even deciding to continue the current activity. Often, monitoring is used in metacognitive control, such as in the earlier example where Kendall used the outcome of her self-tests (monitoring) to decide which planet names to continue studying (control).

The Role of Metacognition in Learning

Metacognition can contribute to student successes and failures. Students who do not accurately monitor their progress may not spend enough time on activities that they believe they are doing well on but in fact have not been learned well. For instance, if a student believes that he is able to solve algebra equations for a particular variable, he may stop practicing this particular kind of problem even if he still requires further practice to master it. Some students may not monitor their progress, which could limit the effectiveness of problem solving. Alan Schoenfeld had novice and expert mathematicians solve a variety of difficult math problems. The major difference between the two was that the experts continually monitored their progress; if they monitored that one approach was not working, they tried another one. The novices failed to monitor altogether, and hence they spent too much time exploring a single—and usually invalid—solution to a problem.

Given that students must monitor (and do so accurately) to efficiently control their study, a common focus of research on metacognition has been to establish how accurately students can monitor their ongoing progress. Unfortunately, across many domains, students have demonstrated limited skill at accurately monitoring their progress. They are relatively poor at judging how well they have learned to associate pairs of words (e.g., cheval—horse); they are relatively poor at evaluating how well they comprehend text materials and often fail to detect inconsistencies in texts that they are reading; they are relatively poor at evaluating how close they are to finding the correct solution for insight problems; and they often have difficulties monitoring the relative effectiveness of a variety of learning strategies. Even though students’ monitoring is often unimpressive, they still use monitoring to make decisions on
how to study. Put differently, students may use inaccurate monitoring to control study, which can lead to poor student outcomes. In fact, almost every teacher has had students perform poorly on an exam, yet afterward, the students complain that they were sure they knew all the material. In these situations, they likely did not accurately evaluate their learning and prematurely stopped studying.

Fortunately, techniques are being discovered to help students accurately monitor their progress in many domains. Kendall evaluated her learning by first studying each planet and its name and then asking herself, “Can I recall the name from memory?” Students who use this technique right after they study each to-be-learned item show poor accuracy at evaluating which items they have learned well, because the memory for each item would be easily accessed from short-term memory. By contrast, students who first study items and then wait to assess their learning show very high levels of accuracy, because by delaying their monitoring they can assess whether each item has been stored in long-term memory. Thus, simply having students delay their monitoring after study can greatly enhance its accuracy. Although delayed monitoring can help students monitor their learning of simple materials (e.g., associating a picture of Saturn with its name), delayed monitoring per se does not always improve accuracy. For instance, students’ evaluation of how well they comprehend text materials does not improve if these evaluations are delayed after reading the texts. Instead, other techniques—such as summarizing each text—may be necessary for achieving accurate evaluations of comprehension.

The Role of Self-Efficacy in Learning

Beyond monitoring-and-control processes, students’ beliefs about their own abilities appear to have an influential role in their classroom performance. In our opening scenario, Kendall believed she could successfully learn science materials, and this high self-efficacy will motivate her to persist while learning even difficult science concepts. More generally, academic self-efficacy refers to a student’s belief about how well he or she can perform in the classroom. To measure efficacy, students rate themselves on statements such as “I know how to study to perform well on tests.” As compared to students with low efficacy, students with higher academic self-efficacy are expected to persist longer and use better strategies to study and hence should have higher academic achievement. As expected, academic self-efficacy of college students is related to cumulative grade point average, even after controlling for other relevant factors such as academic skills and commitment to college.

Relative to research on cognition, research on metacognition is in its infancy. Even so, over 2 decades of metacognitive research involving education has demonstrated the vital role of accurate metacognition to student performance across many disciplines, including memorizing key concepts in the classroom, problem solving, writing, reading, and mathematics.

Keith W. Thiede and John Dunlosky

See also Implicit Memory; Mental Effort

Further Readings

Metaphor

Metaphors—figurative uses of language in which a word or phrase that ordinarily designates one thing is used to designate another—occur across languages and across uses of language, from ordinary conversation to literary and scientific writings. In spite of the prevalence of metaphors, it has proved surprisingly difficult to characterize precisely how they work. Consider a few examples: “Juliet is the sun,” “My surgeon is a butcher,” “Vanity is the quicksand of reason.” Each of these metaphors brings together otherwise unrelated entities to achieve informative and perhaps insightful effects in a remarkably compact way. Comprehending them requires the hearer (or reader) to draw on the literal meanings of the words used, but broader knowledge about the entities literally denoted by the words in question—the
sun, butchers, and quicksand—also seems to be required. This discussion will focus on two questions about metaphor that have received particular attention from philosophers, linguists, and psychologists in recent years. The first question concerns the nature of metaphorical content: What, if anything, does a metaphor mean beyond the literal meanings of its words? The second question concerns how hearers construct metaphorical interpretations: Do they treat metaphors as implicit comparisons or as explicit categorizations?

**Metaphorical Content**

Although metaphors are notoriously resistant to being paraphrased, there is no doubt that they can provoke extensive and vivid effects. The nature and status of these effects has been a matter of considerable recent debate. Four main positions have been defended regarding the nature of metaphorical content and how that content is related to the literal meanings of the words (or sentence) uttered: a broadly semantic account, a non-cognitivist account, and two pragmatic accounts, one in terms of implicature, the other in terms of direct content.

According to semantic accounts of metaphor, particular words or phrases of the metaphorical sentence are reinterpreted so that the sentence as a whole takes on a new (metaphorical) meaning. Whereas early versions of the semantic account attributed this to an interaction or conflict between the literal meanings of the sentence’s subparts, more recent work by Josef Stern posits a metaphor-operator in the sentence’s logical form that demands metaphorical instantiation. The metaphor-operator is a covert marker in the underlying structure of the sentence, attaching to the part of the sentence that is interpreted metaphorically and directing the interpretive process toward the generation of a metaphorical, rather than literal, interpretation.

Although semantic accounts give full credit to the centrality and importance of metaphorical language, it remains difficult to defend the idea that metaphor is a matter of semantic, that is, linguistically encoded, meaning, rather than of language use. Early semantic accounts struggled to explain the status of metaphorical meanings in relation to literal meanings; insofar as metaphorical meanings are both novel and contextually sensitive, it’s difficult to see why they should count as an aspect of what the words of the metaphor mean, rather than something that the speaker means (which she manages to convey by uttering those words). Stern’s metaphor-operator neatly sidesteps this worry, but at the cost of a substantial increase in the complexity of the general process of semantic interpretation.

Partly in response to worries about early semantic accounts, philosophers such as Donald Davidson have moved to the other extreme, denying that metaphors have any determinate content beyond their literal meaning. Although agreeing that a given metaphor may give rise to a wide range of thoughts (or other effects) in its hearer, this “non-cognitivist” view insists that a metaphor does not express those thoughts. Instead, the metaphor simply causes its hearer to notice a range of thoughts by nudging her to see one thing as another (e.g., to see a certain surgeon as a butcher). The non-cognitivist view thus allows for the full variety of metaphors’ effects, its capacity to stir emotions and stimulate vivid mental images or a sense of new insight. Further, the difficulty of paraphrasing metaphors is easily explained: If a metaphor does not have a distinctive metaphorical content, then it is unsurprising that attempts to capture this putative content routinely fail. Nonetheless, the view has been widely criticized for denying that there is anything to get right or wrong when interpreting a metaphor, for we routinely take speakers who utter metaphors to be making truth-evaluable claims and we agree and disagree about both the substance and correctness of those claims.

Intermediate between the semantic and non-cognitivist extremes are two pragmatic accounts of metaphor. Both accounts endorse the idea that metaphors communicate a determinate content but reject the idea that the sentence itself has a metaphorical meaning in addition to its literal one. Instead, metaphorical content is something that the speaker communicates by speaking as she does: It is what she means, not what her words mean. These positions are distinguished by whether they take the metaphorical content to be communicated to the hearer directly or indirectly, that is, to constitute something that the speaker asserts or something that she merely implies.

On the indirect account, deriving from the work of H. P. Grice and John Searle, the speaker is taken to be saying (or “making as if to say”) exactly what her words literally express. Insofar as this literal interpretation is inadequate as a contribution to
the ongoing conversation (e.g., it is patently false or irrelevant), the hearer is led to consider the possibility that the speaker is making a more oblique conversational contribution. If someone says, “My lawyer is a shark,” she probably isn’t intending to assert that her legal representative is a large fish; instead, she should be interpreted as intending to communicate indirectly, describing her lawyer as a shark in order to convey that the lawyer has various qualities typically associated with sharks: perhaps that he is ruthless, vicious, or predatory.

One strength of the indirect pragmatic account is that it shows how metaphorical content might depend on literal meaning without needing to posit metaphorical meanings as well. However, metaphors seem to function much more directly than this view allows: It is perfectly acceptable to answer Romeo’s claim that Juliet is the sun by saying “Indeed she is” or “No, she isn’t,” and these are responses to a claim that Romeo has asserted, not to something he has merely implied. Further, the claim that metaphorical utterances are processed by first computing the sentence’s literal meaning has been challenged on empirical grounds. Work by Sam Glucksberg and colleagues has shown that when subjects are asked to judge the literal falsity of sentences, they take longer to reach a decision when the sentence has a plausible metaphorical interpretation. This suggests that people have difficulty ignoring metaphorical interpretations, and more specifically, that those interpretations do not depend on the inappropriate-ness of a literal alternative.

The final option is to take metaphorical content to be communicated directly by the speaker, that is, to be something that the speaker asserts by her utterance, by allowing asserted content to extend well beyond sentence meaning. One cognitively oriented example of this position is developed within the framework of relevance theory by Robyn Carston, Dan Sperber, and Deirdre Wilson. According to this account, there is no presumption that speakers aim to use words literally, but only that their utterances will be relevant enough to their hearers to be worth processing, and as relevant as the speakers themselves are willing and able to make them. The hearer thus undertakes to construct an optimally relevant interpretation, employing a single interpretive mechanism in both metaphorical and literal cases. Encoded concepts are replaced as needed by “ad hoc” concepts with broader or narrower denotations in order to generate an interpretation of the speaker’s utterance able to support a range of implications sufficient to achieve optimal relevance. For example, suppose that a speaker utters the sentence “John is a soldier” to say of John (someone who is not literally a soldier) that he is metaphorically a soldier—perhaps that he is steadfast, loyal, and can be counted on in a tight spot. Whereas the encoded concept soldier picks out the set of soldiers, the interpretive process draws on contextually relevant assumptions associated with the encoded concept, such as being steadfast and loyal, to construct a related, ad hoc concept with a broader denotation than that of the encoded concept—a denotation that picks out not just actual soldiers but also other individuals (such as John) who share these associated properties of soldiers. This ad hoc concept is then taken to capture the property the speaker is asserting of John.

This account respects both the directness of metaphorical utterances and the ubiquity and naturalness of their occurrence in ordinary discourse. At the same time, the direct pragmatic account is consistent with psycholinguistic work on both the automaticity and relative effortfulness of metaphor processing. One might ask, though, whether the account reduces the difference between metaphorical and literal cases too much: The interpretation of especially poetic metaphors seems to involve more conscious reflection and explicit appeal to the metaphor’s ordinary literal interpretation than the process of ad hoc concept construction looks to allow.

Metaphor Comprehension

Arising in counterpoint to questions about the nature of metaphorical content are questions about how hearers construct metaphorical interpretations. When a metaphor juxtaposes two entities, what sort of connection is the metaphor thereby presenting? Specifically, is the metaphor comparing the two concepts in question, or is the metaphor’s topic being categorized as belonging to a group of which the metaphor’s vehicle is a typical or exemplary member? (A metaphor’s “topic” is its subject, while its “vehicle” is that thing the topic is compared to or described as being—e.g., Juliet and the sun, respectively.)

Comparison models begin with the idea that metaphors involve feature matching, that they
highlight similarities between the metaphor’s topic and vehicle. However, there are a number of difficulties with any straightforward feature-matching account. Metaphors routinely highlight correspondences between features that are not literally shared by topic and vehicle. For example, the sun is a source of literal warmth to anyone who feels its rays, but Juliet is not—she is, at most, a source of emotional warmth to Romeo. Further, good metaphors tend to involve new and informative ways of thinking about their topics by allowing us to see how certain properties of the vehicle might be carried over to the topic.

Dedre Gentner and colleagues have developed a more nuanced comparison model based on the idea of “structure mapping.” This account models the comprehension process in two steps: First, an alignment process looks for structural correspondences between the two conceptual domains of the metaphor (e.g., the domain of lawyers and the domain of sharks). These correspondences may be indirect, involving relational properties that, although instantiated differently, play similar roles within the two domains. So, although we might say that lawyers and sharks are both predatory, in one case the predation relation is entirely social while in the other it’s rather more carnivorous! Further, the model prioritizes matching of interconnected systems of properties over piecemeal matching of individual features. The second step involves projecting unmatched properties within the aligned system of concepts from the vehicle to the topic (e.g., the property of being the center of the solar system is mapped to the property of being at the center of Romeo’s life and thoughts). The structure-mapping account thus aims to generate structural and relational resemblances beyond direct feature matching while using projection to explain a metaphor’s informativeness.

However, this account does not explain how so-called emergent properties are derived. These are properties not standardly possessed by either topic or vehicle, yet which are central to the metaphorical interpretation. For example, “That surgeon is a butcher” communicates that the surgeon is careless, but carelessness is not a property typically associated with either surgeons or butchers. Further, the structure-mapping account does not explain a striking discrepancy between metaphorical and literal comparisons: Only the former can be transformed into meaningful class-inclusion statements. Whereas “My job is a jail,” literal comparisons such as “Emeralds are like sapphires” become nonsense when turned into the categorical “Emeralds are sapphires,” suggesting that something other than comparison underlies the metaphorical case.

Sam Glucksberg has taken this discrepancy to show that metaphors really are what they appear to be: categorical statements, rather than implicit comparisons. According to his “category-transfer” account, comprehending a metaphor requires using the metaphor’s vehicle to construct a superordinate or abstract category of which that vehicle is a prototypical member—the metaphor asserts that its topic belongs to this category. In the metaphor “My lawyer is a shark,” the word shark refers not simply to the set of sharks but to the set of predatory creatures more generally—it has dual reference. However, because a single vehicle may be used metaphorically to convey more than one meaning (we would mean something quite different were we to say that Achilles, rather than Juliet, is the sun, for example), the metaphor’s topic must play a filtering or constraining role in selecting the correct metaphoric category in a given instance.

Assigning the topic a merely “filtering” role may not be sufficient, though: It relegates the topic to a passive role in the interpretive process and might entail unnecessary processing effort by requiring that alternative metaphorical categories be computed. Further, as with the structure-mapping account, the category-transfer account lacks a clear explanation of emergent properties. More generally, the two accounts seem best suited to explaining different kinds of metaphors: Structure mapping looks most plausible for novel metaphors, whereas category transfer more naturally explains more conventionalized cases.

Empirical and theoretical investigations of these questions are increasingly proceeding hand in hand. Both developmental and clinical populations are proving rich sources of data for evaluating the competing models of metaphor discussed here. At the same time, new methods of collecting data, such as neuroimaging, are also being brought to bear, further expanding the range of available evidence.

*Catherine Wearing*

*See also* Analogical Mapping and Reasoning; Conversation and Dialogue; Discourse Processing, Models of; Inferences in Language Comprehension
Microgenesis of Consciousness

The purpose of this entry is to suggest how microgenetic theory, a fundamentally new paradigm for understanding the relation between brain and mind, can be used to illuminate one of the oldest and most fundamental problems in psychology and the philosophy of mind—the nature of consciousness. The entry begins with an explanation of microgeny and the structure of the mental state as such (a perception, an action, a feeling), then proceeds to demonstrate how consciousness arises in the transition from one state to another.

The Structure of the Mental State

Microgeny is the process by which a momentary mental state is formed across successive, qualitatively different phases that represent (in a rather literal sense of that term) phases in brain evolution (phylogeny), whereas ontogeny is replicated in the processual aspects of the microgenetic sequence. The clinical observation of cognitive, emotional, and behavioral symptoms in brain-damaged patients provides evidence in support of a process-based approach to the brain/mind problem. Microgenetic theory is contrary in both substance and spirit to much contemporary theorizing in psychology and the neurosciences, still largely dominated by cognitivism and information theory, but it harks back to certain ideas found in older theories from psychology (William James, Sigmund Freud, Gestalt psychology), behavioral neurology (John Hughlings Jackson, Karl Goldstein, Alexander Romanovich Luria) and philosophy of mind (Charles Peirce, Henri Bergson, Alfred North Whitehead).

Microgenetic theory offers a theory of the mind/brain state as such, that is, without reducing mind to brain or brain to mind. Mapping the cortex actually tells us very little about what the brain is, or what it does, because the brain is a four-dimensional object that does what it does precisely by changing and becoming. Microgeny begins with the “reptilian” brain, that is, the brain stem and hypothalamus, the first part of the central nervous system to appear in phylogeny (similar to Plato’s thymos and Freud’s Id), then moves to the “paleomammalian” brain (the limbic system, the seat of emotion and some aspects of memory), and from there to the cortex. The overall movement, literally and metaphorically, is from an inner core outward to the periphery, from simple wholes to the increasingly detailed parts, where the outside world is not the beginning of the process, but the end.

From a philosophical standpoint, the theory postulates that mental or external objects are not solid or static entities, such as the solid chair out there in the world, but that there is a brief microtemporal history in the mind that is part of their structure. In its journey out from the mind to the world, the chair passes through unconscious stages of form, concept, and meaning relations, in which the figural appearance of the chair, its recognition and category-relations to other similar objects, and to the life experience, are traversed. This means that the chair is not the solid object it appears to be, but is the outcome of a dynamic series of phases.

Consciousness is interpreted in microgenetic theory, then, as the relation across these phases in the mental state, that is, the relation of the empirical self (the subjective center of experience) to images in personal space and/or objects created by the mind/brain in an external space that is the final phase of a subjective process of object and space formation. Specifically, it is a relation across the phases that constitute a single mind/brain state, an epoch of microgenetic time, as depicted in Figure 1.
Thus, conscious perception is not a passive process of recording what is actually “out there,” Kant’s *Dinge an sich*, but rather it places the self in relation to objects that emerge from within, located in a specially objective space and time grid that is created by the brain/mind, constrained by sensory data only in the last stages of the process to fit with reality. The unconscious self operates with a limited repertoire of instinctive pre-objects and behaviors oriented to survival with no concept of past or future, or any space other than the immediate perceptible field, or any objects other than those recognized there as belonging to a limited set of primitive categories, or any possible actions other than stereotyped reactions to objects or pre-objects that conform to these categories. The limbic system (a cluster of gray-matter nuclei that lie above the brain stem and under the neocortex) imbibes objects with the feelings they arouse, pain or pleasure, delight or disgust, amusement or boredom. The neocortex appears last in phylogeny and is more highly developed in humans than the higher primates (especially the frontal lobes). The neocortex mediates the final phase of object- and action-formation. The developing configurations undergo a final adaptation to the environment through the influence of sensory constraints. The instinctual activity of the core can be contained within the microgenetic epoch, which lasts for milliseconds, but feeling states can be more or less persistent (revived), and thoughts can be maintained (recur more or less continuously) for a lifetime. In microgenetic theory, the duration of persistence is a function of the iteration or recurrence of the object, while its revival in memory is a recurrence in the present.

Microgeny is an adaptive becoming in which the environment, in the form of sensation, trims or parses away what is maladaptive to the inferred physical surround but does not provide the stuff of which perception is made. Rather, the brain “thinks up” the world it perceives. The perceptible world is the furthermost rim of mind in a rapid transition from inner core to outer surface. The physical world is known indirectly through its model in consciousness; its nature depends on the type of organism and the adequacy of sensory data, but consciousness cannot sustain the world without the constraints of sensation. The degree of approximation to an objective world determines the content of the state, for example, daydream, reverie, fantasy, hallucination, delusion, confabulation, and so forth, all of which are psychic events that actualize at different points in object formation. Without the impact of sensation, especially at the final phase of the microgeny, thinking is dreaming; psychosis is an intermediate phase.

In this “bottom-up” sequence, sense data do not enter the act/object development, but remain external, where they sculpt or constrain what develops. Sense data, even within the sculpting process, are imperceptible: We perceive the effects of sensory parsing in perception but not the assembly of sensory bits. In the transition from limbic memory to neocortical perception, there is a cascade of whole-part or context-item shifts that leads from inner and private to outer and public, from events that are memory-like to those that are perception-like, from past to present, from concept and image to external object, from the archaic to the recent in forebrain anatomy, from unity to diversity, and from the simultaneity of the unconscious to temporal order in conscious recollection and in the world. The sequence recurs in overlapping fashion every fraction of a second. The recursive nature of this process causes the stream of consciousness to appear as overlapping moments in an unsteady, even occasionally chaotic rhythm of repetition and novelty.

**Subject and Self, Awareness and Consciousness**

A subject is the subjective whole of the organism excluding its external portion. The external portion of that whole is made up of objects, perceived as...
existing outside the organism and belonging to the physical world. Objects are psychic appearances, literally phenomena ("that which appears" in ancient Greek), which, on reflection, point to entities that putatively exist independently of what we think we know of them. Awareness refers to the relation of subject to object. The conscious self, a segment in the stream of outgoing subjectivity aware of its own priority and subjectivity, arises within a subject that is aware of objects. Consciousness can be conceived as consisting in the relation of a self to inner and outer objects, where the relation is the unidirectional process of becoming through which subject, then object, actualizes. The relation of the self to inner objects is introspection or reflection; to outer objects, exteroception or perception.

The core, an early phase in the epoch, is the seed of personality. In higher mammals and young children, the core shows the first tendency to individuality, expressed in temperament or attachments. Prior to its appearance, behavior seems to be regulated by mechanism: instinctual drive, environmental signal, consummation. The core or unconscious mind is dependent on drive and the immediate occasion. In the evolution of mammals, individuality and awareness show a gradual advance from species to species, as the core becomes more detached from the object world, capable of entering into the binary "I/Thou" relation out of which all subsequent interpersonal relations evolve. Out of the core a conscious self is shaped in relation to beliefs and values and guided to actuality by sense data. The self is conscious of inner and outer objects, fixed in the present by the attention system and short-term memory, yet able to attend to events in the past (long-term, episodic, and semantic memory) and to plan for those in the future (prospective memory, imagination).

As partition continues, object and lexical concepts, images and feelings, punctuate the subjective pole. At the objective pole, value penetrates objects with greater specificity and refinement. The simple archaic category of "edible" is broken down into "tasty" or "not tasty" (essentially "I like it" or "I don't like it"). Conditioning later aesthetic judgments of increasing subtlety, while requiring the intervention of a subject, the "I" who does or does not like. There is a parallel articulation of inner and outer. The present encompasses events of greater duration, enfolding a narrative of self and experience. The bodily space of the core expands to the perimeter of the arm's reach, that is, a "manipulation" space, then beyond this perimeter to an external world that finally "detaches" from the observer. The action space of young infants transforms to the independent space of the conscious adult. We see an analogous transition from the implicit and unconscious to the conscious and explicit in all areas, for example, a word that individuates from a "mental lexicon," a specific recollection from a "memory store." The transition from concept to object, store to item, lexicon to word, unconscious to conscious, is not a transfer of like to like, as if the depth were a mere container. The transition of category to instance or whole to part occurs over a qualitative series of largely hidden internal phases that constrain consciousness to an outcome that delimits a theoretically infinite number of possibilities.

### Conclusion

William James regarded consciousness as the central issue in psychology and the duration of the present as the central issue for its understanding. Time and space are specified out of the core: Space as objects grow out of concepts, time as duration is incremented by events. Subjective time develops in a transition from the simultaneity of the core to serial order in the world. Inwardly, time is counted in duration, externally in increments. The duration required for events to be perceived as stable objects is also the basis for the perception of events. The persistence of an object over a minimal duration to be perceived for what it is entails a recurrence within successive nows. All objects are events in which change (recurrence) is imperceptible.

Duration is the "glue" of continuity that carves events out of flux. Time is not a uniform flow but a replacement of changing objects across intervals, themselves changeless, thus nonexistent. The continuity of the self, of inner and outer, and the recognition of sameness or difference, owes to the overlap in a succession of present moments. Specifically, the overlap of the present (now) in the replacement of a categorical self and its objects is the basis for the near-identity of recurrences. The scenario of incessant change with a relative stability of inner and outer events is comprehensible in terms of categories sufficiently flexible to accommodate deviance and sufficiently habitual to cancel brief atypical replications.

Maria Pachalska,
Bruce Duncan MacQueen,
and Jason Walter Brown
See also Conscience, Comparative Perspectives; Consciousness and Embodiment; Consciousness and the Unconscious; Self-Consciousness; Time Perception

Further Readings


MIND-BODY PROBLEM

The mind-body problem is the problem of explaining how the mind and body are related. Put that way, the problem seems singular even if not simple. Is the mind simply the body or some part of it? If it is not, then what is it and how is it related to the body? Most important, if the mind is not the body, how do they interact?

Descartes and Dualism

The most obvious strategy for solving the problem is to insist that minds are bodies or some proper part of the body, perhaps the brain or the central nervous system. This position, the identity theory, claims that minds and brains are identical. René Descartes famously argued that the identity theory must be false. According to Descartes, it is conceivable that minds exist without bodies, and so it is possible that minds exist without bodies, and so minds are not bodies. Now the relationship between what is conceivable and what is possible is tricky and has attracted much attention. It is dubious that Descartes is licensed to move from conceivability to possibility. But what of the move from the notion that it is possible that minds are not bodies to the conclusion that minds are not bodies? That move might seem a non-starter. Much that is possibly the case is not actually the case. But what Descartes assumes, at least implicitly, is that if minds and brains are identical, then they must share all of their properties in common. Descartes assumes that whatever is identical must be indiscernible: For any two objects, x and y, if x is identical to y, then for any property x has y has, and vice versa. And so if minds are brains, then whatever property the one has the other also has. But it is not possible for brains to exist without brains. And so, if minds are brains, then minds could not exist without brains. Thus, if it is possible for minds to exist without brains, then it follows that minds are not brains.

Descartes' response to the mind-body problem is that minds and bodies are radically different kinds of things; the latter material, the former immaterial. But Descartes' solution, dualism, introduces two new problems. The first challenge is to make intelligible what an immaterial thing might be. The second problem, related to the first, is the problem of mental causation: How do minds causally interact with bodies? This problem for Descartes is especially acute: How do immaterial substances interact with material substances?

The contemporary version of the mind-body problem is not cast in terms of substances and is often cast, instead, in terms of laws, states, properties, or events. In terms of properties, for instance, the mind-body problem is this: How are mental properties related to physical properties? That is, how are mental properties related to the properties of interest to the natural sciences? And the problem of mental causation, in its modern guise, concerns the causal closure of the physical. It is commonly presumed that the laws of physics are causally closed; there are no physical events that are not caused by other physical events. Were this not the case, then there would be, at the explanatory level of physics, miracles: physical events with no physical explanations. But
if the instantiations of mental properties are causally efficacious with respect to the physical, if having some mental property is the cause of some behavior, for instance, then we are left with only three possibilities. First, there are physical events that are not physically caused, and so there are physical events without physical causes. Second, whatever the mental causes the physical also causes, and so behavior is causally overdetermined. Third, mental properties are identical to physical properties. The first option is the modern version of Descartes’ solution. The third is the modern version of materialism, what is commonly called “physicalism.” The second? Well, about that there is much contemporary debate, to which we will return.

**Eliminativism**

A more radical way to respond to the problem of mental causation, however, is simply to deny that the mental is causally efficacious. One way to do this is to claim that the mental, although it might be caused by the nonmental (e.g., the feeling of pain might be caused by a broken bone) and although it might cause what is mental (e.g., someone’s experiencing pain might cause that person to remember being in pain), does not cause anything nonmental (and so the feeling of pain would not cause someone to avoid the cause of that pain in the future). This position, epiphenomenalism, has few contemporary adherents. The problem for the position is that, first, it seems obvious to many that the mental is causally efficacious, and second, if the mental does not cause the physical (if, e.g., it does not cause behavior), it is hard to see why we should believe it to exist.

Early in the 20th century, the more popular response to dualism’s troubles was to deny the mental entirely, or at least to deny it any explanatory role. According to psychological behaviorism, what is explanatorily relevant is what is observable, and what is observable with respect to explaining human behavior are environmental stimuli and behavioral outputs. Psychological behaviorism attempts to “solve” the mind-body problem by denying the mental entirely. Another form of behaviorism, logical behaviorism, agrees with psychological behaviorism that what matters is what is observable, but insists that the mental is causally efficacious, or at least that appeals to the mental play an explanatory role. Logical behaviorists attempted to define mental terms in terms of behaviors or dispositions to behave. By comparison, we might think of something’s being fragile as its being disposed to shatter if struck. If it shatters when struck, we might say that it shattered partly due to its fragility. But its fragility, it might be thought, just is its being such that it would shatter. There is an explanation, logical behaviorism contends, even though the explanatory circle is small. Logical behaviorism does not deny the mental, per se. Ordinary psychological explanations have a purchase. But, according to logical behaviorism, the mental is not separate from the body. Indeed, logical behaviorism denies that the mental is at all private.

In philosophy, the popularity of both versions of behaviorism went hand in hand with a theory of meaning that was popular during the early part of the 20th century, the verificationist theory of meaning, according to which the meanings of theoretical terms were supposedly the verification conditions for their correct applications. Verificationism and the behaviorist theories following in its train now have few adherents. With better theories of meaning and reference, and especially with better theories with respect to theoretical terms, both philosophy and the behavioral sciences once again began appealing to hidden objects, states, and processes without embarrassment.

Not everyone agreed that these included the mental, however. The latter part of the 20th century introduced new versions of eliminativism. According to defenders of this view, attributions of propositional attitudes (e.g., beliefs and desires) belong to a folk theory for explaining behavior. As a theory, this folk theory is subject to revision and even elimination as better theories are developed. If the best theories for explaining human behavior appeal to neuroscience, and if our folk theories are not reducible to the theories of neuroscience, then, according to these eliminativists, we should conclude that our folk theories of the mind are not simply incomplete, but false.

**Functionalism**

Another strategy that surfaced during the second half of the 20th century, and that has attracted far more attention, is functionalism. According to this view, someone is in pain, for instance, whenever she is in a state that plays a particular functional role, which is to say whenever she is in that state
commonly caused by pain stimuli (e.g., pin pricks) and that commonly produces pain behavior (e.g., moving one’s arm quickly away from the pin). Functionalism shares with behaviorism the idea that the meanings of mental predicates are tied with behavior, but it breaks with behaviorism in that, according to functionalism, mental predicates pick out properties and states that not only cause behavior but also cause each other. Pains produce not only behavior but also beliefs and memories. And part of the commonsense explanation for this appeals to the fact that we reflect on our memories, beliefs, and desires and that this reflection results in additional memories, beliefs, and desires.

One version of functionalism, realizer functionalism, is generally thought of as a version of the identity theory. According to realizer functionalism, because mental terms pick out whatever it is that plays a particular functional role, and because it is presumably the job of neuroscience to tell us what that is, mental state terms (at least when used to ascribe mental states to us) pick out neurophysiological states (of us). Thus understood, realizer functionalism solves the problem of mental causation by insisting that mental properties just are certain physical properties, and so there is no threat to the causal closure of the physical and there is no worry about causal overdetermination.

Another version of functionalism, role functionalism, holds that mental terms pick out properties and states at a higher level of abstraction, perhaps those that humans might share with machines having no neurophysiology. As commonly understood, role functionalism holds that mental properties are second-order properties—properties had by properties. Someone is in pain, according to this view, in virtue of having some first-order property (presumably some physical property) the instantiation of which plays the appropriate functional-causal role. So, for instance, if having some neurophysiological property plays the pain role for us, if having that neurophysiological property is typically caused by pin pricks, for example, and typically causes pain behavior, then we are in pain whenever that neurophysiological property is instantiated in us. But that neurophysiological property is not identical to pain. Rather, pain is the having of some property or other (in our case that particular neurophysiological property) that plays the relevant functional-causal role.

Role functionalism is theoretically compatible with both dualism and physicalism about the mind. It is compatible with dualism since it allows that the instantiation of some nonphysical property might play the appropriate functional role. It is compatible with physicalism since, presumably, nothing nonphysical actually plays that role. According to role functionalism, the property of being in pain is not, strictly speaking, a physical property or any particular physical property, and so creatures physically very different than us, and perhaps even machines of the right type, might share the property of being in pain. But assuming that pain is always realized physically, assuming that everything that is in pain has some physical property the instantiation of which plays the functional role of pain for that creature or machine, there is an important sense in which dualism is false.

Exactly how and whether dualism is false and physicalism is true by the lights of role functionalism has itself been an area of intense debate. According to one particularly influential view, token-token physicalism, although mental types (e.g., pain) are not identical to physical or neurophysiological types, particular instances or tokens of the mental (e.g., someone’s particular pain at some particular moment) are identical to particular instances or tokens of the physical. By analogy, no one thinks that a particular shade of red, say crimson, is identical to redness, because something can be red but not be crimson. Nonetheless, it might be held, something’s being crimson is just its being red.

The Mind-Body Problem Today

Role functionalism was once thought to be a kind of philosophical panacea, a halfway house between a dualism that seems shrouded in mystery and a reductive physicalism that seems counter to common sense, and it continues to be the favored view in the philosophy of mind. But role functionalism faces several objections. According to one objection, being functionally organized in a particular way is insufficient for intentionality. We can imagine, it is thought, a creature or a robot with the right kind of functional organization, but without what we would think of as thoughts or understanding.

It is also challenging to see how functional organization is sufficient for consciousness, but here the
problem is seemingly a problem for any theory of the mind, and especially for any physicalist theory. It seems possible, for instance, that someone functionally like me, someone who typically sees blue objects as blue, who typically comes to believe that such objects are blue after seeing them, who typically asserts that such objects are blue when asked, and so forth, might have experiences of blue objects that are qualitatively different than the experiences that I have when I see blue objects. We are functionally the same, she and I, but we are not mentally the same. The problem, the qualia problem, is to explain how functionalism in particular, and physicalism more generally, can explain the qualitative nature of many of our mental states.

Where the qualia problem raises the worry that functionalism cannot avoid all of the problems of physicalism, another problem suggests that role functionalism cannot avoid a perennial problem for dualism. The problem, the problem of mental causation, is that if mental properties are not identical to physical properties, then it would seem that their instantiations would have nothing to do. One version of the problem maintains that any causal explanation of an event excludes all other explanations. The problem, if such there is, would seemingly apply to the properties appealed to by most of the natural sciences. Indeed, it would seem to raise a problem for all of the sciences except for physics since, if the argument were successful, it follows that the lower level causal explanations exclude all other higher level causal explanations. But role functionalism seems to face special challenges in explaining how the mental could, by its lights, be causally efficacious. If a mental property is a second-order property, if it is the having of some property or other that plays a particular functional role, then it would seem to be the first-order property that does the work by definition. And if instances of mental properties are identical to instances of physical properties, if the role functionalist endorses token-token identity, then those instances must play exactly the same causal role. And if they play the same causal role, if instantiations of mental properties never cause anything not caused by instantiations of physical properties, then that would seem a good reason to think either that mental properties do not exist or that they just are those physical properties.

One particularly interesting response to this problem is to think of mental property tokens not as identical to physical tokens, but instead to think of them as proper parts of physical tokens. The position agrees with realize functionalism that mental properties are not second-order. And it agrees with dualism, but here a dualism about properties and not substances, that instances of mental properties are not identical to instances of physical properties. It also agrees with role functionalism that the mental is physical in virtue of being multiply realizable by the physical. To understand the suggestion, imagine an overly simplified view, a popular philosophical fiction. Imagine that pain in humans is correlated with C-fibers firing and that pain in Martians is correlated with A-Fibers firing. According to traditional role functionalism, if John the human is in pain, then his being in pain is not his having C-fibers fire, although it is true that he wouldn’t be in pain if his C-fibers were not firing. His being in pain is his being functionally organized in the appropriate way and, thus, his having something or other that is playing the pain role. Perhaps we might say that his being in pain at this moment is identical to his having C-fibers firing at this moment, but being in pain is not having C-fibers fire. But then it is unclear what causal role there is for pain to play for John since any instance of pain in John just is an instance of C-fibers firing in John.

According to traditional realize functionalism, if C-fibers firing is the realizer of pain for humans, then just is pain for humans. And because A-fibers firing just is pain for Martians, it just is pain for them. But then it appears that humans and Martians, although it is true to say of both that they experience pain, actually share nothing in common when they are both in pain.

If we think that pain is multiply realizable, if we agree with the Role Functionalist that humans and Martians have something in common when in pain, then we must also think that the causal powers of C-fibers firing and the causal powers of A-fibers firing are relevantly similar. No doubt they do not share all of their causal features, for otherwise they would not be neurophysiologically distinct, but their causal features must overlap with respect to those features relevant for pain, for otherwise we would not treat them as alike psychologically. Pain, on this view, contributes the set of causal powers that is a
proper subset of those powers contributed by both C-fibers firing and A-fibers firing, and so an instance of pain might be thought of as a proper part of any instance of C-fibers (or A-fibers) firing. Whether this strategy will prove successful is as yet a matter of considerable controversy.

Michael Watkins

See also Anomalous Monism; Behaviorism; Consciousness and Embodiment; Eliminative Materialism; Emergence; Explanatory Gap; Mental Causation; Physicalism; Reductive Physicalism

Further Readings

MIRROR NEURONS

Our social competence largely depends on the capacity to understand the intentional behavior of others. What are the origins of this capacity? What are its underlying neural mechanisms? This entry will present and discuss a class of neurons originally discovered in the premotor cortex of macaque monkeys that can shed light on these issues: mirror neurons.

Mirror Neurons in Monkeys

In the early 1990s, a new class of premotor neurons, “mirror neurons,” was discovered in the anterior sector of the macaque monkey’s ventral premotor cortex, known as area F5. Mirror neurons discharge not only when the monkey executes goal-related hand motor acts such as grasping objects, but also when it observes other individuals (monkeys or humans) executing similar motor acts. Neurons with similar properties were subsequently discovered in regions of the posterior parietal cortex reciprocally connected with area F5.

Action observation causes in the observer the automatic activation of the same neural mechanism triggered by action execution. For the first time, a neural mechanism allowing for a direct matching between the visual perception of an action and its execution has been identified. By means of the mirror matching mechanism, the results of the visual analysis of the observed action—which, in principle, has no meaning for the observer—can be translated into an account that the individual is able to understand. It was proposed that this mechanism could underlie a direct form of action understanding. If mirror neurons do in fact mediate action understanding, their activity should reflect the meaning of the observed action, not its visual features.

Two sets of experiments were carried out to verify this hypothesis. The first experiments tested whether the mental representation of an action triggers F5 mirror neurons, the second whether mirror neurons can respond to the sound produced by actions. The results of these experiments answered both questions in the affirmative and showed that what drives mirror neurons’ discharge is not the mere visual description of a motor act, but rather its goal.

In the most lateral part of area F5, a class of mirror neurons responding to the execution and observation of mouth actions has been found. The majority of these neurons discharge when the monkey executes and observes mouth-related, object-related motor acts, such as grasping, biting, or licking. However, a small percentage of mouth-related mirror neurons discharge during the observation of communicative facial actions performed by the experimenter in front of the monkey (“communicative mirror neurons”). Macaque monkeys show an initial capacity to control and “voluntarily” emit social signals; this is mediated by the frontal lobe. It is interesting that this capacity develops in a cortical area—area F5—that in humans became Brodmann’s area 44, a key area for verbal communication.

More recently, premotor and parietal mirror neurons have been found to have a role in intention
understanding. The discharge of mirror neurons during the observation of an act (e.g., grasping an object) is conditioned by the type of subsequent act (e.g., bringing the object to the mouth) that specifies the overall action intention. In addition to recognizing the goal of the observed motor act, mirror neurons allow the observing monkey to predict the agent’s next action, and hence its overall basic motor intention. This neural mechanism could provide scaffolding for more sophisticated social cognitive abilities, such as those that characterize the human species.

**Mirror Neuron Mechanisms in Humans**

Solid evidence, recently also at the single neuron level, demonstrates the existence of a mirror neuron matching mechanism in the human brain as well. Action observation leads to the activation of premotor and posterior parietal cortical areas, the likely human homologue of the monkey areas in which mirror neurons were originally described. Distinct cortical regions within the premotor and posterior parietal cortices are activated by the observation/execution of mouth-, hand-, and foot-related actions.

The mirror neuron mechanism for actions in humans is directly involved in imitation of simple movements and in the imitative learning of complex skills. Furthermore, many interesting phenomena described by social psychologists, such as the “chameleon effect”—the unconscious mimicry by the observer of postures, expressions, and behaviors of her or his social partners—can find a neurophysiological explanation in the mirror mechanism. The premotor cortex, which has the mirror mechanism for action, is also involved in processing action-related words and sentences, suggesting that mirror neurons, along with other parts of the sensory-motor system, could play a role in language processing.

Mirroring mechanisms also underpin our capacity to empathize. When we perceive others expressing a particular emotion such as disgust, the same brain areas are activated as when we experience the same emotion. This, of course, does not imply emotional contagion. In fact, in spite of a common shared activation focus of activation in the anterior insula, no matter whose disgust is at stake, different cortical areas activate when disgust is subjectively experienced as opposed to when it is only observed in the facial expression of someone else. Similar mirror mechanisms have been described for the perception of pain and touch.

Together, these results suggest that our ability to empathize with others is mediated by mechanisms of embodied simulation, that is, by the activation of the same neural circuits that underpin our own emotional and sensory experiences. In this view, empathy is conceived of as the outcome of our natural tendency to experience interpersonal relations at the implicit level of intercorporeity, that is, of the mutual resonance of intentionally meaningful sensory-motor behaviors. Recent studies have suggested that mirror mechanisms could be malfunctioning in individuals affected by autistic spectrum disorders. The discovery of mirror neurons opens exciting new perspectives in a variety of different fields in social cognitive neuroscience, such as our understanding of psychopathological states, language, and aesthetics.

The mainstream view in cognitive science was, and may partly still be, that action, perception, and cognition are to be considered as separate domains. The discovery of mirror neurons challenges this view by showing that such domains are intimately intertwined. It also provides a new empirically based notion of intersubjectivity, which can be viewed first and foremost as intercorporeity, the main basic source of knowledge we directly gather about others.

*Vittorio Gallese*

**See also** Common Coding; Consciousness and Embodiment; Emotional Recognition, Neuropsychology of; Facial Expressions, Emotional; Motor System, Development of; Social Cognition

**Further Readings**


Mnemonic Strategies

This entry reviews mnemonics (memory improvement strategies), describing effective techniques that have been developed over 2,500 years but which have been shown to make a very considerable contribution to the ease of learning disconnected or relatively meaningless material.

Imagery Mnemonics: The Method of Loci

Mnemonics are strategies to improve memory through contrived associations. Their use dates back to classical times. Cicero, writing in 55 BCE, described how, more than 400 years earlier, the Greek poet Simonides discovered the method of loci. Simonides, as the only survivor from a banquet at which the roof collapsed, was able to identify the bodies of the victims by recalling images of the seating arrangements. He realized that location was a powerful cue; if he formed mental images incorporating an image of each item to be remembered with each of several familiar, ordered locations, then the items could be easily recalled by reimagining those locations. From classical times onward, this method of loci (place) was used by orators to remember important points in long speeches, and modern research has confirmed the method’s effectiveness at doubling recall.

The method of loci illustrates key features of successful mnemonic methods. The familiar locations provide reliable cues to recall and the images allow the incorporation of otherwise unrelated items into a familiar framework. A mental trip through the locations allows each to cue the image and help recall of the items. Normally, we remember familiar, meaningful, organized, and interesting information without special techniques, but faced with needing to learn apparently meaningless, disorganized, disconnected, and uninteresting information, mnemonic methods help by providing the meaning, organization, and retrieval cues that are lacking.

Imagery Mnemonics: Peg Words

The method of loci has obvious limitations: It requires a known set of ordered locations and is particularly appropriate for serial recall. To create greater flexibility, professional memory experts developed alternative methods that use images to link items to previously memorized peg words, with a separate peg word for each of 100 or more items. For example, the peg word for the 21st item to be remembered might be net and for the 22nd might be nun. Here, images involving a net or a nun replace the location of the method of loci, and knowing that the 21st peg is net cues recall of the relevant image. Different professionals advocate different systems for constructing the pegs, but a commonly used method translates each digit into a consonant sound and then concrete peg words are constructed by inserting vowels. So, in the phonetic mnemonic system, 1 translates as a t sound and 2 as an n sound—from which the examples of net and nun were constructed. This translation system not only allows the construction of peg words that can be easily memorized but also enables numbers to be translated into memorable phrases, so aiding, for example, the learning of PIN numbers or dates. (The remaining digits are 3 = m, 4 = r, 5 = 1, 6 = j/sh, 7 = k, 8 = f/v, 9 = b/p, 0 = s/z.)

The Linkword Mnemonic

The keyword or linkword mnemonic is a further development of imagery-based mnemonics that has been applied very successfully to the learning of foreign language vocabulary. The sound of each foreign word is converted into an easily imaged word or phrase that sounds similar. The image is linked to the meaning of the word. For example, the Spanish word perro sounds like pear and means dog. So an image of a dog eating a pear would cue the meaning (dog) when the word perro was encountered.

Imagery mnemonics have been adopted, with varying results, to aid those suffering memory problems caused by brain damage. The benefits are inversely proportional to the degree of deficit. In general, the traditional mnemonic methods are very effective but demand effort and creativity on the part of the learner.

The effectiveness of imagery-based mnemonics derives from the opportunity that the images provide for spatially linking known and to-be-recalled items together when other ways of associating the to-be-recalled items with memory cues may be lacking. Merely forming an image of something to be remembered does not enhance long-term recall. It is only when images are formed that integrate the known (i.e., peg or loci) with the unknown (item to be recalled) that later recall is improved. Any technique that provides cues and a meaningful
framework can improve memory. Linking the items to be remembered in a story provides such a meaningful framework and can dramatically improve recall.

**Acronyms and Rhymes**

Other mnemonics popular with students include acronyms, in which the first letters of words to be remembered are arranged to spell out a meaningful word, and acrostics, in which the first letters of the words become the first letters of words in an easily memorized phrase. For example, the acronym FACE has often been used to remember the notes in the spaces of the treble clef and the acrostic “Every Good Boy Deserves Favor” cues the notes on the treble clef lines. Research has shown such mnemonics can be effective, but primarily as a way of remembering the order of known items rather than for memorizing new information. Rhymes, such as the “Thirty days hath September” rhyme for the lengths of the months, are another technique, and rhymes and rhythm have been shown to help by limiting the possible alternatives during recall.

**Applying and Combining Mnemonics**

Mnemonics are most effective when the learner combines ability and commitment; when mnemonophobia (fear of using mnemonics) is overcome, they are effective in classroom learning. However, other memory improvement techniques incorporating retrieval practice, such as the name game for learning the names of members of a group, are less demanding on the learner and can be equally effective. Retrieval practice and imagery mnemonics can be combined to achieve even greater memory improvement than when either is used separately. Suitable selection of techniques from among mnemonic and other memory improvement techniques can dramatically ease the burden of memorizing and the embarrassment of memory failure.

*Peter Morris*

**See also** Memory Recall, Dynamics; Retrieval Practice (Testing) Effect; Visual Imagery

**Further Readings**


**MODELING CAUSAL LEARNING**

Humans display remarkable ability to acquire causal knowledge. Hume’s philosophical analysis of causation set the agenda for discovering how causal relations can be inferred from observable data, including temporal order and covariations among events. Recent computational modeling work on causal learning has made extensive use of formalisms based on directed causal graphs (Figure 1). Within a causal graph, each arrow connects a node representing a cause to an effect node, reflecting core assumptions that a cause precedes its effect and has some power to generate or prevent it. The computational goal is to infer from the observable data the unobservable causal structure conveyed by the graph and the magnitude of the power of each cause to influence its effect.

This entry covers some key issues that arise in modeling human causal learning. Alternative models of causal learning vary depending on the assumptions adopted in the computation, the goal of the computation, and the presentation format of the input data. Understanding models from these perspectives can clarify their commonalities and differences, guide the design of psychological experiments to test the validity of key assumptions, and assess whether models can potentially be extended to

![Figure 1](image_url)
real-life problems, such as medical diagnosis and scientific discovery.

Alternative Causal Assumptions

When the causal graph includes multiple potential causes of a single effect, two leading classes of models make different assumptions about the integration rule used to combine causal influences. One class (including the classic delta-P model and the associative Rescorla-Wagner model) assumes a linear integration rule: Each candidate cause changes the probability of the effect by a constant amount regardless of the presence or absence of other causes. A second class is represented by the power PC theory, a theory of causal judgments postulating that learners assume that unobservable causal influences operate independently to produce the effect. Guided by this assumption, causal integration is based on probabilistic versions of various logical operators, such as OR and AND-NOT, chosen to reflect the polarity of causal influence (i.e., whether a cause produces or prevents the occurrence of an effect).

When the causal graph includes multiple effects of a single cause (e.g., flu causes headache and chest pain, as shown in Figure 1), the causal Markov assumption states that the probability of one effect occurring is independent of the probability of other effects occurring, given that its own direct causes are present. Statistical models that examine causal relationships adopt the Markov assumption, which guides exploration of conditional independencies that hold in a body of data, thereby constraining the search space by eliminating highly unlikely cause-effect relations. The extent to which humans employ the causal Markov assumption remains controversial.

When observations are limited, human causal learning relies heavily on some type of prior knowledge. Prior knowledge can be specific to a domain (based on known categories), but it can also include abstract assumptions about properties of a system of cause-effect relations (e.g., preference for causal networks that exhibit various types of simplicity). Use of appropriate prior knowledge can explain the rapid acquisition of causal relations often exhibited by humans.

Alternative Types of Causal Judgments

Causal learning potentially enables two types of judgments: causal strength and causal structure. A strength judgment involves a quantitative assessment of a cause-effect relation: What is the probability with which a cause produces (or prevents) an effect? A structure judgment is a more qualitative assessment of cause-effect relations: Does a candidate cause, in fact, produce (or prevent) an effect? In a causal graph, a strength estimate assesses the weight associated with an arrow, whereas a structure estimate assesses the existence of an arrow. Learning causal strength allows a system to anticipate the occurrence of an effect in a new context and respond adaptively. Learning causal structure allows a system to explain events and to choose appropriate actions.

A general model of causal learning needs to achieve both of these computational goals. Although many heuristic models of strength estimation have been proposed, most do not generalize to structure judgments. Within a Bayesian framework, a strength-learning model assumes the learner has some prior guesses about possible values of causal strength and updates beliefs given observed data. A structure-learning model assumes the learner has prior guesses about strength and structure, and computes a score to quantify the support that observed data provide for beliefs about aspects of the structure. Bayes’s rule provides the computational engine to update beliefs. A different variant, constraint-based models of structure learning, operates in a more bottom-up fashion, computing statistical dependencies in the data and selecting those structures consistent with these dependencies.

Alternative Learning Formats

Formal modeling of causal learning has emphasized acquisition of causal relations from complete statistical data (i.e., a 2 × 2 contingency table tallying frequencies of the four possible combinations of presence versus absence of cause and effect). Some models restrict themselves to predictions based on such idealized data presentation, side-stepping issues related to memory and presentation order.

In many realistic situations, however, such summary data are not available to the learner. Rather, data about cause-effect pairings arrive sequentially with no external record of the events. For nonverbal animals, there is no obvious way to present summarized data; humans also must often learn from sequential data. Because sequential models aim to
Moral Development

Moral Development

Moral development dynamically integrates prior beliefs with new observations in a trial-by-trial manner, such models are sensitive to order of data presentation. The development of sequential models of causal learning is currently an active research area.

Hongjing Lu

See also Concepts, Development of; Knowledge Acquisition in Development; Reinforcement Learning, Psychological Perspectives

Further Readings


MORAL DEVELOPMENT

Morality is a controversial topic and a source of debate in philosophy and psychology. One source of argumentation is due to different perspectives on whether morality is a matter of mind or heart or, to put it differently, reason or emotion. From the point of view of how morality develops in individuals, debates often center on whether children incorporate teachings from adults (representing societal values and standards) or construct ways of thinking about social relationships that entail understandings of issues pertaining to welfare, justice, and rights. Corresponding debates are over whether morality is a matter of habits or processes of reasoning.

The debates over morality parallel debates in psychology, other social sciences, and philosophy regarding the role of reasoning and its connections with emotions in human functioning. On one side are those who presume that, in most realms, people act out of habit or non-rationally. On the other side are approaches presuming that thought and reasoning are central. As succinctly put by the philosopher Martha Nussbaum (1999), “human beings are above all reasoning beings” (p. 71). To say that humans are reasoning beings doesn’t, by any means, exclude emotions. Indeed, Nussbaum sees close links in that emotions do not stand alone or overwhelm thought, but are guided by ways of judging social relationships, are part of people’s goals in life, and inform their understandings of other people and events. In such a framework, moral judgments include an integration of reasoning and emotions such as sympathy, empathy, compassion, and respect.

This entry focuses on approaches that have examined the development of moral reasoning, as integrated with emotions. First, the entry provides a historical overview describing early and influential theory and research on the development of moral judgments. It then describes research that modified those early theories through findings that young children construct moral judgments about welfare, justice, and rights, which they distinguish from the customs and conventions of society. Finally, the entry considers how moral and other types of social thought are involved in decision making.

Early Research on the Development of Moral Judgments

Jean Piaget presented in 1932 one of the first extensive analyses of the development of morality from such a framework in his classic work, The Moral Judgment of the Child. Piaget studied several dimensions of children’s thinking about rules and justice. He studied children’s judgments about the rules of marble games, lying, causing material damage, and punishments for wrongdoing. Piaget proposed that moral judgments constitute ways of thinking that take one form in early childhood and are transformed into another form of thinking by late childhood and adolescence. Specifically, he proposed that the development of moral judgments proceeds from a heteronomous to an autonomous level. At the level of heteronomous thinking, children are unable to take the perspectives of others and think in literal ways about rules and authority. They view rules as fixed and sacred and authority as requiring obedience. Heteronomy also involves an inability to distinguish the social from the physical, as well as an inability to take into account intentions or internal psychological states; hence, they judge by consequences rather than intentions. Emotions of fear, sympathy, and respect contribute to the
heteronomous way of thinking. In particular, there is what Piaget referred to as unilateral respect for adult authority. A central feature of heteronomous thinking is that children do not distinguish or differentiate moral ideas from adherence to existing rules and customs or from obedience to authority.

According to Piaget, the shift to autonomous thinking involves differentiations of moral concepts pertaining to welfare and justice from existing customs, rules, and adherence to the commands of authorities. For this shift to occur, it is necessary for children to engage more in peer interactions and to focus less on adult-child interactions. Peer relationships promote both a sense of equality and abilities to understand the perspectives of others. At the autonomous level, children develop their own understandings of the purposes of rules and the roles of authority in social relationships entailing fairness, rights, and cooperation. Children are then better able to take the perspectives of others and distinguish moral aims based on welfare and justice from existing rules and customs. Unilateral respect is replaced by mutual respect by which there is concern with reciprocity in social relationships.

Autonomy does not mean that people are individualistic or concerned with the self or independence. Rather, it refers to participation in generating understandings of moral concepts that are applied to decisions as to how people should treat one another. One of the major subsequent contributions to this type of perspective came from Lawrence Kohlberg, who included children, adolescents, and young adults in his research. He proposed modifications in Piaget's propositions through formulations of the development of moral judgments that included six stages grouped into three levels. At the first level (labeled premoral), children are described as making moral judgments first on the basis of punishment (stage 1) and then on the basis of self-interest and prudence (stage 2). At the next level (conventional morality), moral concepts are not differentiated, first, from maintaining good relations and approval of others (stage 3) and then from the rules, conventions, and authority of social systems (stage 4). It is at the third level of postconventional morality, which presumably does not emerge until late adolescence or adulthood, that moral understandings of justice and rights are clearly differentiated from premoral and conventional understandings; moral judgments are based on concepts of social contract and agreement on procedures in application of laws (stage 5) and on moral principles of welfare, justice, and rights separate from particular social systems (stage 6).

Piaget and Kohlberg made some lasting contributions to the study of moral development. They charted a view of the psychology of morality that included substantive definitions of the moral domain. Many philosophers, including Immanuel Kant, John Stuart Mill, John Rawls, Ronald Dworkin, and Martha Nussbaum, have presented analytic formulations of concepts of justice (see especially John Rawls's influential treatise, *A Theory of Justice*), welfare, and rights. Piaget and Kohlberg demonstrated that the study of the psychology of morality could not be adequately conducted in the absence of sound definitions of this domain. They were, therefore, guided in their psychological research by philosophical conceptions as possibly involving the formation of substantive moral concepts (as opposed to research that only looked at psychological variables like learning, internalization of values, and personality traits).

**Morality as Distinct From Other Domains**

The work of Piaget and Kohlberg also demonstrated that children could be productively posed with problems regarding how people should relate to each other. However, subsequent research shows that the levels they proposed do not accurately capture children's moral judgments because the research tasks used were overly complex. Kohlberg, for instance, presented children with situations entailing conflicts with a number of features to consider. A well-known situation is the story of man who must decide whether to steal a drug he cannot afford to save the life of his wife who has cancer. The situation includes several components in conflict, including the value of life, property rights, law, and obligations in personal relationships. Other research has examined the judgments of children, adolescents, and adults about straightforward issues bearing on physical harm, psychological harm, theft, and fairness. This body of research, conducted by Elliot Turiel, Larry Nucci, Judith Smetana, Melanie Killen, Cecilia Wainryb, and Charles Helwig, first involved studies aimed at determining the criteria that children apply to moral and non-moral issues. In a large number of studies, children were presented with situations describing a
transgression identified as part of the moral domain, such as one child physically assaulting another. They evaluated the acts and answered questions ascertaining if they considered the evaluation of the acts to be based on rules (e.g., what if there were no rule about it?), authority (e.g., what if the teacher said it was all right?), and common practice (e.g., what if it was generally accepted in a group or culture?). By 3 or 4 years of age, children’s judgments about moral issues are not contingent on rules, authority dictates, or accepted practices. Acts that harm others or involve unequal treatment are evaluated as wrong even if there are no rules or an authority deems them acceptable. In addition, moral prescriptions are judged to generalize across social contexts. These judgments are based on understandings of welfare, justice, and rights. The research shows, however, that young children’s concepts are primarily about harm or welfare, whereas older children develop concepts of justice and rights along with concerns with welfare. The development of moral judgments also involves increased capacities to relate concepts of welfare, justice, and rights to other considerations in complex situations.

It is well established that young children differentiate morality from punishment, obedience, authority, and interests of the self. Children distinguish the domain of morality from the domain of social conventions. Social conventions refer to existing regularities in social systems that coordinate social interactions (examples are customary practices about matters like forms of address, modes of dress, eating habits). At all ages, conventions are judged to be contingent on rules, authority, and common practices. Moreover, children form ways of thinking about the personal domain, which pertains to arenas of personal jurisdiction that do not involve impinging on the welfare or rights of others.

Children develop distinct ways of thinking in the moral, conventional, and personal domains. The domains constitute different developmental pathways, with age-related changes within each domain. Observational studies in homes and schools also show that social interactions around moral events differ from interactions around conventional events. Interactions associated with moral events typically do not involve communications about rules or expectations of adults (which do occur for conventional events), but are about feelings and the perspectives of those involved, as well as communications about harm and fairness. The development of moral judgments is associated with the early emergence of emotions like sympathy, empathy, and respect.

**Moral and Social Decision Making**

The existence of different domains of social reasoning has implications for explanations of how people make decisions in social situations. In coming to decisions, people take into account different domains, different considerations within the moral domain, and different priorities. Two examples—honesty and rights—illustrate the process of coordinating different considerations in coming to decisions. Honesty is generally considered morally right and necessary to maintain trust. However, in some circumstances, honesty is not necessarily the morally correct course of action. An example discussed in philosophical discourse is a situation in which lying might be necessary to save a life. In this regard, it has been found that physicians judge deception of medical insurance companies acceptable when it is the only means to obtain treatment for a patient with a serious condition. They give priority to preventing harm over honesty. Other studies have shown that adolescents consider it acceptable to defy and deceive parents who direct their offspring to engage in acts considered morally wrong. Deception of parents is also judged acceptable when they direct activities seen as part of adolescents’ legitimate personal choices. Similar results of the coordination of honesty and moral or personal considerations have been obtained in research on adults’ judgments regarding marital relationships. Similarly, coordination of different considerations is seen in decisions about rights like freedom of speech and religion. Although children and adults endorse rights, in many situations they subordinate rights to matters like preventing harm or promoting community interests.

To understand moral and social decision making, it is necessary to examine the different domains of thought that people apply to social situations. Therefore, the study of moral development requires examination of social and personal domains, as well as the moral domain.

Elliot Turiel

See also Concepts, Development of; Emotion and Moral Judgment; Folk Psychology; Knowledge Acquisition in Development; Social Cognition
Motivated Thinking

Once controversial, the notion that people's motivations can influence their thoughts now features prominently within many areas of psychology and plays an important role in current research on memory, reasoning, decision making, and perception. The effects of motivation on cognition can be conceptualized as stemming from three general sources: (a) motivations to use particular types of judgment strategies (e.g., a focus on minimizing missed opportunities versus eliminating mistakes); (b) motivations to achieve broad, content-independent (nondirectional) types of judgment outcomes (e.g., decisions as concise and unambiguous, or as accurate as possible); and (c) motivations to achieve narrow, content-dependent (directional) types of judgment outcomes (e.g., impressions of oneself as successful or loved). Whereas motivations for judgment strategies primarily affect the quality of cognitive processing that occurs, and motivations for nondirectional judgment outcomes primarily affect the quantity of processing, motivations for directional judgment outcomes often affect both the quality and quantity of processing. Thus, in addition to being “cognitive misers” whose biases result from generally limited cognitive processing capacity, people are also “motivated tacticians” whose biases result from specific changes in cognitive processing that serve their current goals.

Strategy-Motivated Thinking

Motivations for particular judgment strategies can arise from many different concerns, but those most thoroughly examined relate to concerns with attaining growth (promotion) or maintaining security (prevention). Promotion motivations produce gain-oriented strategies focused on achieving advancement, whereas prevention motivations produce loss-oriented strategies focused on maintaining a satisfactory state. Promotion motivations thus elicit inclusive modes of cognitive processing to identify opportunities for gain, whereas prevention motivations elicit exclusive modes of cognitive processing to minimize losses. For example, when promotion-focused, people consider a broader variety of explanations during causal reasoning, engage in more creative and divergent thinking during problem solving, and attend more to abstract and global properties of a stimulus. In contrast, when prevention-focused, people consider a narrower selection of causal explanations, engage in more analytical and convergent thinking, and attend more to concrete and local stimulus properties. Thus, motivated judgment strategies can influence the quality of cognitive processing that occurs across many domains.

Nondirectional Outcome-Motivated Thinking

Beyond motivations to use particular judgment strategies people may also have motivations to reach particular judgment outcomes. Some types of outcome motivations have been labeled nondirectional because they do not involve specific desired conclusions and focus on more general objectives during judgment. The two most-studied nondirectional outcome motivations are desires for accuracy and desires for closure (conciseness, clarity). Because these desires do not concern the specific contents of a judgment, they primarily affect the quantity rather than the quality of cognitive processing that occurs. Whereas desires for accuracy increase how many explanations people consider during causal reasoning, the effort they dedicate to evidence evaluation and information search, and how much information they retrieve from memory, desires for closure have the opposite effect. Accordingly, judgment complexity increases with desires for accuracy and decreases with desires for closure, whereas simple reliance on recently or frequently activated knowledge during judgment increases with desires for closure and decreases with desires for accuracy. However, these processing differences do not always result in more valid conclusions when motivated by accuracy.
Because of limitations in cognitive resources or access to necessary information, biases can remain even when accuracy motivation is active. Thus, desires for accuracy or closure affect the quantity of cognitive processing during judgment more than they affect how good a judgment is made.

Directional Outcome-Motivated Thinking

Other types of motivations for particular judgment outcomes have been labeled directional because they do specify particular desired conclusions. Directional outcome motivations include people’s desires to believe they are competent, socially connected, and in control. Whatever the desired conclusion, directional outcome motivations affect cognitive processing in two ways: They alter the quality of cognitive processing in ways that selectively highlight evidence for this conclusion, or they alter the quantity of cognitive processing such that evidence for this conclusion is accepted after only a cursory review but evidence opposing this conclusion is thoroughly scrutinized.

Some examples of quality-related effects are that desires for perceptions of competence encourage explanations that accept responsibility for success but not failure, encoding and activation of knowledge that emphasizes potential for success, and selecting standards of comparison that imply higher levels of ability. Similarly, desires for perceptions of social connection encourage explanations that diminish relationship conflict, encoding and activation of knowledge that emphasizes commonalities with valued others, and selecting standards of comparison that imply higher levels of belonging. Some examples of quantity-related effects are that desires for perceptions of competence encourage more thorough processing of competence-threatening feedback and generation of more hypotheses undermining such feedback. Similarly, desires for perceptions of personal control encourage an extended analysis of evidence that contradicts individuals’ basic understanding of how the world works.

Although common and extensive, effects of directional outcome motivations do have limits. Whatever their motives, people still recognize they must meet standards of objectivity while forming judgments. Thus, directional outcome motivations have the strongest influence when the evidence is uncertain or ambiguous, and typically function to intensify judgments that support prior beliefs rather than produce new conclusions.

Influences of motivation on cognition are pervasive, and studying these influences provides important insights into the human mind. Having settled the first-generation question of “Does motivated thinking occur?” research on strategic, nondirectional, and directional motivation is now considering second-generation questions of when and how effects of motivations on cognitive processing arise. Answers to such questions will provide a deeper understanding of the motivation-cognition interface and further advance the field of cognitive science.

Daniel C. Molden and E. Tory Higgins

See also Cognitive Dissonance; Debiasing; Decision Making and Reward, Computational Perspectives; Placebo Effect

Further Readings


Motor Learning, Practical Aspects

Motor skills are an essential part of our lives. From toddlers attempting to walk, children throwing and catching balls, young adults learning to ski, to older adults or those with physical disabilities trying to regain walking and balance capabilities—throughout our lifetime, we learn and perform motor skills. Even though motor skills vary widely in type and complexity, the learning process that individuals go through when acquiring various motor skills is similar. During the first phase (so-called cognitive stage), considerable cognitive activity and attentional capacity is required, and movements tend to be controlled in a relatively conscious manner. The result of using conscious control strategies is that the movement is relatively slow, fragmented, and inefficient and that the outcome is rather inconsistent. The second phase of learning...
Motor Learning, Practical Aspects

(associative stage) is characterized by subtle movement adjustments. The movement outcome is more reliable, and the performance is more consistent from trial to trial. Inefficient muscular co-contractions are gradually reduced, and the movement becomes more economical. In addition, at least parts of the movement are controlled more automatically. After extensive practice, the performer reaches the autonomous stage, which is characterized by fluent and seemingly effortless motions. Movements are accurate, with few or no errors, consistent, and efficient. The skill is performed largely automatically at this stage, and movement execution requires little or no attention.

How can the learning process be facilitated and individuals' ability to perform or maintain those skills be enhanced? This is a question that interests practitioners (e.g., coaches, physical therapists, athletes, musicians) and theorists alike. Studies have identified a number of factors that influence learning. Yet, the functioning of those factors has been viewed mainly from an information processing perspective. Only recently has it become clear that the learning process is not merely the acquisition of a specific movement pattern that is facilitated by providing learners with the right information at the right time. Learning also encompasses affective reactions, the self-regulation of cognitive processes, and attentional focus to meet task demands. This is particularly relevant in the natural and almost inevitably social context of movement. Thus, both the learner's informational and motivational needs have to be optimized to enhance learning.

To help learners acquire the goal movement pattern and reach a state of automaticity, they are typically provided with demonstrations, instructions, and feedback. The following sections discuss how the effectiveness of these learning variables can be enhanced by taking into account both their informational and motivational roles. When considering learning, it is important to keep in mind that learning is assumed to reflect a relatively permanent change in a person's capability to perform motor skills. Therefore, in experimental studies, learning is typically assessed in retention or transfer tests (the latter involve a variation of the task), performed under the same conditions for all groups and at a given time interval following practice under different conditions of interest.

Demonstrations

Demonstrations (e.g., live or video presentations of a model) are often used in practical settings to provide the learner with an idea of the goal movement. Studies have shown that learning through observation is effective, especially if it is combined with physical practice. Observational practice appears to be particularly beneficial for the learning of complex motor skills, and combined observational and physical practice can be more effective than physical practice alone. It has been argued that observation gives learners the unique opportunity to extract important information concerning appropriate coordination patterns or subtle task requirements, or to evaluate the effectiveness of strategies—which would be difficult or impossible to do while executing the movement. From that perspective, observational practice offers the learner a chance to conduct information processing that could not occur while physically practicing.

Demonstrations can involve expert models or another learner (learning model). Observing a learning model can be as effective as observing an expert, particularly when two learners practice in a dyad (pair) and alternate between physical and observational practice. In some studies, dyad practice resulted in more effective learning than individual (physical) practice—even though dyad participants had only half the number of physical practice trials that participants in the individual, physical practice group received. Aside from the information gained by observing another learner, learning benefits of dyad practice are presumably also a result of enhanced motivation, resulting perhaps from competition with the partner, the setting of higher goals, or the loss of self-consciousness as people fulfill interdependent dyadic roles and find another in the same learning boat. It is perhaps not coincidental that participants in collaborative or cooperative learning situations often anecdotally report more enjoyment than they have experienced learning alone.

Thus, interspersing physical practice with demonstrations can make an important contribution to skill learning. Considering the high costs of certain types of training (e.g., pilot training, medical education, physical therapy), the incorporation of demonstrations, dyadic, or collaborative practice, may not only be cost-efficient but can also enhance the effectiveness of training.

Instructions

Focus of Attention

Studies have shown that the wording of instructions has an important impact on performance and
learning. Specifically, instructions directing attention to performers’ movements (inducing a so-called internal focus) are relatively ineffective. In contrast, directing attention to the effect of their movements on the environment, such as on an implement (inducing an external focus), generally results in more effective performance and learning. For instance, focusing on the swing of a golf club has been demonstrated to lead to greater shot accuracy than focusing on the swing of one’s arms. On tasks involving balance, focusing on the movements of the support surface results in greater stability than focusing on the movement of one’s feet. The learning advantages of instructions promoting an external focus have been shown for a variety of motor skills, levels of expertise, and populations (including children and persons with motor impairments).

An external focus of attention appears to speed up the learning process, or shorten the first stages of learning, facilitating movement automaticity. In contrast, a focus on one’s own movements results in a more conscious type of control that constrains the motor system and disrupts automatic control processes (constrained action hypothesis). Movement efficiency has also been shown to be enhanced by an external focus. The mere mention within the internal focus instructions of the performer’s body may act to increase self-consciousness, or self-focus, which, in turn, may lead to self-evaluation and activate implicit or explicit self-regulatory processes in attempts to manage thoughts and affective responses.

Conceptions of Ability

Individuals’ beliefs about, or conceptions of, key abilities have been shown to affect the learning of motor skills. Specifically, whether people view their ability as something that is genetically determined (i.e., reflecting a fixed capacity or “talent”) versus something that is amenable to change with practice influences their motivation and learning. Even though most adults have certain ability conceptions, these can also be influenced by instructions given in a learning situation. Some researchers have manipulated those conceptions to assess their influence on individuals’ motivation and performance of motor skills. In these studies, ability concepts were induced through instructions depicting performance on the task as something that reflected either an inherent ability or an acquirable skill. Learners who viewed the task as an acquirable skill, as opposed to reflecting an inherent capacity, showed greater self-efficacy, more positive affective self-reactions, greater interest in the task, as well as more effective learning and greater automaticity in movement control.

The construction of a task as something that reveals one’s inherent capacity may act as a threat to one’s ego. Learners who view a task as a reflection of an inherent ability presumably approach the learning situation with more apprehension than those who see task performance as an acquirable skill. This, in turn, may hinder the learning process compared with a situation that is regarded by the performer simply as a learning opportunity. Similar to other variables (e.g., attentional focus and normative feedback), a person’s ability conception appears to affect the extent to which he or she becomes self-conscious—with concomitant effects on motor performance and learning.

Feedback

Frequency of Feedback

Views regarding the role of feedback (knowledge of results, knowledge of performance) in motor learning have changed considerably over the past century. Whereas early researchers believed that feedback should be given frequently and immediately after the movement, this notion changed in 1984 when Alan Salmoni, Richard Schmidt, and Charles Walter proposed the guidance hypothesis. According to this hypothesis, feedback guides the learner to the correct movement pattern, while at the same time carrying the risk that—if provided frequently, immediately after, or even concurrently with, the movement—the movement—learners might become dependent on it, thereby failing to develop the capability of detecting and correcting errors themselves. In addition, the learning of a stable movement representation has been assumed to be made more difficult by frequent feedback, due to increased variability in performance in the learner’s attempts to correct errors. Numerous subsequent studies have provided support for those assumptions, for example, by showing that reducing the feedback frequency or delaying feedback can be beneficial for learning.

However, recent findings suggest that detrimental effects of frequent or immediate feedback may occur primarily when the feedback induces an internal focus (i.e., directs attention to the body movements). If the feedback promotes an external focus (i.e., directs attention to the desired movement effect), a high frequency has been shown to be more effective
than a reduced feedback frequency—presumably because it helps learners maintain an external focus. Even concurrent feedback can be beneficial for learning if it induces an external focus.

**Self-Controlled Feedback**

Having learners decide after which trials they want, or do not want, to receive feedback has been demonstrated to lead to more effective learning than predetermined feedback schedules. The percentage of practice trials on which self-control learners requested feedback varied widely between studies. Yet, the actual feedback frequency appears less important than the learner’s ability to choose, or not to choose, feedback. Self-controlled practice conditions have generally been assumed to lead to a more active involvement of the learner, enhancing motivation and increasing the effort invested in practice. Self-controlled feedback might also produce better correspondence to learners’ needs for information about their performance, such as after a strategy change, or allow them to ask for feedback after presumably successful (more motivating) trials.

**Positive Feedback**

Feedback after “good” trials has been demonstrated to enhance learning compared to feedback after “poor” trials. In studies in which learners were provided with feedback after blocks of trials, groups who received feedback about their best trials in that block (and not about the worst trials) showed superior learning compared to those who received feedback about their worst trials (and not about their best trials). It is interesting that, in studies on self-controlled feedback, learners indicated that they preferred to receive feedback—and chose feedback—more frequently after relatively successful trials. This may be another reason for the learning benefits of self-controlled feedback.

Similarly, positive or negative normative feedback can affect learning. Normative feedback involves norms such as a peer group’s actual or bogus average performance scores. Thus, normative feedback, by definition, involves social comparison—a ubiquitous phenomenon in settings that involve the performance and learning of readily observable motor skill or lack thereof. In studies in which learners were provided with a (fabricated) average score of other performers, in addition to their own score, it was found that learning was positively or negatively affected depending on whether learners were led to believe that their performance was above or below average. Specifically, the conviction that one’s performance was better than average was associated with more effective skill learning than the belief that one’s performance was below average—essentially resulting in a self-fulfilling prophecy.

Positive or negative performance feedback presumably influences the cognitive perception of personal capability (e.g., self-efficacy expectations, perceived competence) and creates positive or negative affect experienced for the self. Self-efficacy expectations, in turn, may influence individuals’ goal setting, effort, and attention to task performance. Recent neuroscientific evidence links positive affect to the dopamine processing that supports sequence learning. Negative affect may also dampen, or interfere with, memory processing.

Gabriele Wulf and Rebecca Lewthwaite

See also Automaticity; Skill Learning, Enhancement of

Further Readings


**Motor System, Development of**

Motor development involves advances in behavior across the entire body—the eyes and head for looking, the trunk for maintaining a stable postural base, the arms and hands for manual actions, and the limbs for locomotion. Development entails increasing coordination between active muscle forces and passive gravitational and inertial forces. However, there is more to the study of motor development than muscles and biomechanics. Goal-directed movement is inextricably linked with perception, cognition, and social interaction, and motor skill acquisition includes developmental changes in all of
these domains. This entry summarizes four critical aspects of motor development: Movement is ubiquitous, prospective, creative, and malleable.

**Movement Is Ubiquitous**

Movement is the most pervasive and fundamental of all psychological activity. The body is constantly in motion. Some movements occur in the background (breathing, swallowing, postural compensations), some are spontaneous by-products of arousal and brain activity (twitches, shakes, flails), and some are goal-directed (looking, talking, reaching, walking). The massive amounts and variety of children’s motor experiences facilitate discovery of new skills and their improvement. New forms of movement also set the stage for changes in other psychological domains by creating new opportunities to explore the environment and engage in social interactions. Movements in the face and throat make possible behaviors that are fundamental for life, such as sucking, chewing, and swallowing; those required to produce speech are among the most sophisticated movements learned by humans.

The first self-produced movements occur prenatally. Fetuses nod and turn their heads, open and close their jaws, yawn, suck, and swallow amniotic fluid. They wrinkle their foreheads, move their lips and tongue, and, after 25 weeks, open and close their eyes. Whole body movements and large movements of the arms and legs peak at 14 to 16 weeks postconception, then decrease as the growing body fills the uterine space. Some fetal movements are not random: Fetuses direct hand movements toward their own faces and bodies, the wall of the uterus, and the umbilical cord.

The sheer amount of movement is staggering. By 3.5 months of age, infants have performed 3 to 6 million eye movements. By 10 months, infants have accumulated enough crawling steps to travel more than half the length of Manhattan. By 12 months, infants have experienced over 110,000 bouts of wiggles, waves, kicks, and flaps of 47 different types of spontaneous stereotypies. At 14 months of age, infants take about 15,000 walking steps per day. Sleep does not quiet newborns’ active bodies. While sleeping, they stretch, roll, wave, and twitch.

**Actions Are Prospective**

For motor actions to be adaptive, they must be controlled prospectively—guided into the future based on perceptual information about the body and environment. Even the simplest movements of the head and limbs require anticipation of disruptions to a stable postural base. Perceptual feedback from just prior movements informs the consequences of future actions.

Infants’ earliest actions show inklings of prospectivity. At 1 month of age, infants predict the trajectory of a moving target and smoothly follow it with their eyes, but prospective control is fragile and easily disrupted. The target must be large and slow moving or eye movements will lag behind. The development of prospective looking is protracted over several months. By 4 to 5 months of age, predictive looking is sufficiently stable for infants to track targets moving behind an occluder, so that their eyes wait on the far side to spot the target when it reappears.

As with looking, the development of prehension involves increasing prospectivity. In general, infants demonstrate prospective control of looking before reaching, and once they can reach, they frequently bring objects to their eyes for visual inspection. Newborns are highly motivated to keep their hands in view. In a dim room illuminated with only a narrow shaft of light, newborns move their hand when the light beam moves and slow their arm movements before the hand arrives in the light, rather than after their hand appears. Reaching for stationary objects appears at 12 weeks and intercepting moving objects appears at 18 weeks, but infants’ first reaches and catches are jerky and crooked. Infants’ arms speed up, slow down, and change directions several times before the hand finally contacts the toy. After a few months, reaches and catches become more adult-like, with one large movement to bring the hand near the target and a subsequent smaller movement to grasp it. Infants reach for glowing objects in the dark at the same age that they reach in the light, suggesting that they can gauge the location and size of the object and use muscle-joint information about arm position to guide the reach. By 9 months, infants pre-orient their hands to grasp in the dark. By 11 months, infants catch moving objects as they appear from behind an occluder.

Prospective control of locomotion also takes months to develop. When approaching a sheer drop-off or steep slope, novice crawlers and walkers plunge right over the edge. After several weeks of locomotor experience, they guide locomotion prospectively by using perceptual information gathered.
Motor System, Development of

from exploratory looking and touching to decide when cliffs and slopes are safe or risky. With sufficient experience, infants—like adults—can adapt locomotor actions to changes in the environment and in their own bodies and skills. For example, when experimenters load experienced toddlers with lead-weighted shoulder-packs to make their bodies more top-heavy, infants instantly recalibrate their judgments of risky slopes to their new, restricted abilities. They correctly treat the same slopes as risky while wearing lead-filled shoulder-packs and safe while wearing feather-filled shoulder-packs.

The ability to create new possibilities for action with tools also requires prospective control, but before 1 year of age, infants have difficulty planning tool use strategies. For example, 9-month-olds grab a spoon by the bowl-end instead of the handle or hold it with the bowl pointing away from their mouths. They correct grasp errors reactively by switching hands or awkwardly rotating their hand to bring the bowl to their mouth. By 18 months of age, infants know which end of a tool to grasp, how to grasp it, and how to plan their motor actions in advance, but they are still inefficient when using a tool to act on another object (hairbrush on doll) rather than performing an action centered on their own body (hairbrush on self). By 24 months of age, infants prospectively adjust their typical strategies to use tools in a novel way, such as gripping a spoon with a bent handle to scoop food from a bowl.

Solutions Are Creative and Enlist a Variety of Means

The movements of infants (and novices of any age) are notoriously variable and unreliable, whereas movements of adults and experts are smooth and consistent. Over weeks of practice, infants’ visual scanning patterns, reaches, and steps become increasingly efficient, reliable, and predictable. However, all infants do not solve the problem of moving their bodies in the same way. More lethargic infants learn to reach by powering up the muscle forces; more lively infants hone reaching skills by dampening down inertial forces from ongoing arm flaps.

Variety in infants’ spontaneous exploration provides information about objects and surfaces and about the efficacy of the self in control. Spontaneous leg kicks in 2- to 4-month-olds transform into deliberate one-legged, alternating, or simultaneous two-legged kicks as infants explore the contingencies between their movements and the jiggling of an overhead mobile yoked to their legs. By the second half of the first year, infants explore the sound-making properties of objects and surfaces and eventually hang the hard side of objects against the hard side of a tabletop. Visual, manual, and oral exploration are coordinated into bouts of rotating, fingering, and mouthing objects.

Variable routes to development suggest that individual infants explore multiple solutions before settling on the most efficient solution. For example, prior to crawling on hands and knees, infants display a variety of locomotor strategies. They hitch in a sitting position, crab on their backs, and log roll. Belly crawling is so variable that infants change the configuration of limbs used for support and propulsion and the timing between limbs from cycle to cycle. Infants move ipsilateral limbs together like a camel, move contralateral arms and legs together in a near-trot, lift front then back limbs like a bunny hop, and “swim” with all four limbs lifted into the air at once.

When the constraints of infants’ growing bodies and nascent skills preclude adultlike solutions, infants find temporary “placeholder” actions that get the job done. Although 12-month-olds chew well enough to break down food and swallow it, they chew with lateral rather than rotary jaw movements. It takes years before the lips and tongue are involved and cooperating in a planful and deliberate way and before rotary movements are incorporated into the chewing action. Moreover, infants chew the same way for every kind of food, whereas older children flexibly adapt their jaw movements to the food consistency and to the emergence of new teeth and molars. Even habitual actions such as moving the bolus to a consistent “working” side of the mouth take years to develop.

Sometimes infants’ ignorance about conventional motor solutions opens up new means for solving motor dilemmas. When challenged to cross narrow bridges with only a wobbly rubber handrail for support, 16-month-olds use a “light touch” strategy, grazing their hands along the rail to generate somatosensory information for controlling posture, and a “heavy touch” strategy, where they exploit the deformability of the handrail to rappel as if mountain climbing or lean back as if wind surfing. When faced with impossibly steep slopes, infants descend
Multimodal Conversational Systems

Multimodal conversational systems are computer systems that engage human users in intelligent

In the United States, the recent practice of putting infants to sleep on their backs rather than their stomachs has resulted in delayed onset of crawling and other prone skills. In cultures that do not encourage crawling (including American infants circa 1900), large proportions of infants skip crawling altogether, either bum-shuffling or proceeding straight to walking.

Karen E. Adolph and Sarah E. Berger

See also Motor Learning, Practical Aspects

Further Readings


conversation through speech and other modalities such as gesture and gaze. These systems are motivated largely by human-human conversation, where nonverbal communication modalities such as hand gestures, body postures, eye gaze, head movements, and facial expressions are used to complement spoken language. Studies have shown that multimodal conversational systems provide more natural and effective human-machine interaction compared to speech-only systems. This entry provides a brief overview of the types of systems, their general architecture, and key components of automated multimodal interpretation and generation in such systems.

Types of Systems
A variety of multimodal conversational systems have been developed in the past 3 decades. They range from multimodal conversational interfaces to embodied conversational agents and to more recent situated dialogue agents. Multimodal conversational interfaces address interaction with interfaces from computers or other devices (e.g., handheld devices). A user can look at the interface, point to regions on the interface, and talk to the system. These types of interfaces are particularly useful for map-based applications. Embodied conversational agents (also called virtual humans) allow users to carry on conversations with virtual embodied agents (often life-size virtual agents) through multiple modalities such as speech, facial expressions, hand gesture, and head movement. These types of systems are often applied in the domain of cultural training, tutoring, and education. Situated dialogue agents represent a new generation of dialogue agents that are co-present with human partners in a shared world, which could be virtual or physical. In situated dialogue (e.g., human-robot dialogue), the perception of the shared environment and the mobility and embodiment of the partners play an important role in success of the dialogue. In these systems, language processing needs to be combined with vision processing, gesture recognition, and situation modeling. Situated dialogue in virtual worlds can be applied in the domains of interactive games, training, and education, while dialogue in the physical world can benefit a range of applications involving human-robot interaction.

System Architecture
Most multimodal dialogue systems share a similar architecture with four major components: multimodal interpreter, dialogue manager, action manager, and multimodal generator, as shown in Figure 1. The multimodal interpreter is responsible for combining different modalities and identifying the semantic meanings of user multimodal input. Based on the understanding of user intent, the dialogue manager decides what to do in response, for example, ask for clarification or provide information requested by the user. Once this decision has been made, the action manager takes charge of any required backend processes, such as retrieving relevant information. The multimodal generator uses the gathered information to produce specific responses such as multimedia presentations on graphical interfaces or multimodal conversational behaviors for embodied agents. Each of these components is critical to the overall performance of a multimodal conversational system. The multimodal interpreter and generator are the two important components unique to multimodal conversational systems in contrast to traditional spoken dialogue systems.

Multimodal Interpretation
The capability to process and identify semantic meanings from user multimodal inputs is one of the most critical components in multimodal conversational systems. A large body of research has focused on how different modalities are aligned, how different modalities and/or shared visual environments can be integrated to derive an overall semantic representation (such as user intent), and how nonverbal modalities may improve spoken language understanding.

When interpreting human input such as speech, the first step is to automatically recognize what has been communicated (e.g., through speech recognition) and to represent this information as several recognition hypotheses. Previous studies by Sharon Oviatt have shown that using complementary modalities such as speech and pen-based gestures can improve selection of the best recognition hypothesis through mutual disambiguation. Many approaches have been developed for integrating different modalities. For example, rule-based methods and finite state machines have been applied to unify semantic structures from individual modalities based on
multimodal grammar (e.g., grammar that encodes temporal relationships between pen-based gestures and linguistic units). Probabilistic approaches have been used to merge semantic representations from individual modalities based on constrained optimization (e.g., minimizing or maximizing an objective function based on a set of constraints).

Psycholinguistic studies have shown that human eye gaze reflects attention and engagement. Eye gaze is tightly linked with human language processing. Recent advances in eye tracking technology have made it possible to incorporate human eye gaze during human-computer interaction. Studies have shown that incorporating eye gaze in a conversational interface improves automated language understanding at multiple levels, from speech recognition and reference resolution to automated language acquisition.

**Multimodal Generation**

Different types of multimodal conversational systems require different capabilities for multimodal generation. In conversational interfaces, graphical visualization is important to provide better access to, and allow better understanding of, the requested information. Thus, multimodal generation is mostly concerned with how to automatically plan, allocate, coordinate, and present multimedia information, for example, with synchronized graphical and speech output. In systems with embodied conversational agents, automated generation of multimodal behaviors is a major research focus. In addition to natural language generation and speech synthesis, recent work also generates synchronized facial expressions to reflect the different emotional states of an agent, models eye gaze and head nodding to indicate grounding, enables hand gestures to indicate emphasis, and produces shifts in posture to signal the beginning or ending of a conversational turn or segment. The same issues concerning behavior generation are also applied to situated dialogue agents where agents/robots often have physical bodies (e.g., head, face, arms, etc.) which require hardware configurations. To address the challenges of integrating multiple modalities and generating natural interactive behaviors for embodied agents, representation languages (e.g., behavior markup language) are developed to describe behavior elements (e.g., individual modalities such as head, gesture, gaze, speech, etc.) and synchronization of behaviors to control an agent.

In summary, computational models and approaches in multimodal conversational systems are developed based on empirical observations of human-machine interaction and are guided by the cognitive and communicative principles in
Multinomial Modeling

Multinomial modeling is a formal approach to measuring cognitive processes, such as the capacity to store and retrieve items in memory, or to make inferences and logical deductions, or to discriminate and categorize similar stimuli. Although such processes are not directly observable, theoretically they can be assumed to interact in certain ways to determine observable behaviors. The goal of multinomial modeling is to identify which underlying factors are important in a cognitive task, explain how those processes combine to create observable behavior, and then use experimental data to estimate the relative contributions of the different cognitive factors. In this way, multinomial models can be used as tools to measure unobservable cognitive processes.

This entry is organized as follows. First, the type of data used in multinomial modeling is described, and how these data can be used to develop models of this type is explained. A detailed example of multinomial modeling is provided as an illustration. Next, a number of common aspects of multinomial modeling are discussed, including validation testing, models for complex data structures, and the use of multinomial models to test hypotheses about cognitive processing. Finally, current work and future directions for multinomial modeling are outlined.

Data Structure and Model Development

Multinomial models are developed for categorical data, where each participant’s response falls into one and only one of a finite set of observable data categories. These data usually come from a cognitive experiment, where each participant in an experimental group produces a categorical response to each of a series of items; for example, pictures are “recognized” or “not recognized” or letter strings are judged to be “words” or “nonwords.” Most data sets for multinomial modeling involve more than two response categories, and in addition there may be more than one type of item, each with its own system of response categories. For example, in a source-monitoring experiment, participants study a list of items from two sources, Source A or Source B (e.g., presented in a male vs. female voice, or presented visually vs. auditorily). Later, participants are given a recognition memory test consisting of three types of items, namely, the two types of old list items and new distracter items, and they must classify each tested item as Source A, Source B, or New. The resulting multinomial data structure consists of three category systems (A, B, or New), each with three response categories (i.e., participants indicate if each item is from Source A, Source B, or is new). If the responses in different category systems are independent and category counts within a system follow a multinomial distribution, the probability of the data...
structure is given by the product of three multinomial distributions, one for each category system.

To express this more formally, assume that the data structure for an experimental task consists of $J$ categories and $N$ experimental response observations, where $n_j$ observations fall into category $C_j$, $j = 1, 2, \ldots, J$. Then if the observations are independent and identically distributed with probability $p_j$ of falling into category $C_j$, the category count vector, $D = (n_1, n_2, \ldots, n_J)$, follows the multinomial distribution given by

$$
\Pr[D|p = (p_1, \ldots, p_J)] = \frac{N!}{n_1! \cdots n_J!} \prod_{j=1}^{J} p_j^{n_j},
$$

where the category probabilities are nonnegative and sum to one.

The key to creating a multinomial model is to take a multinomial data structure such as the one above and express the category probabilities in terms of underlying, cognitively interpretable parameters. One needs to specify a cognitive processing architecture along with formal computational rules that can generate the count data in terms of the parameters. Once the model is constructed and data are collected, standard tools in statistical inference can be used to analyze the data and evaluate the adequacy of the fit of the model to the data. In addition, one can estimate the values of the cognitive parameters that are likely to have created the data. In this way, unobservable cognitive processes can be measured indirectly with the use of the model.

Multinomial models of various types have been used in cognitive psychology since the 1960s; however, in the 1980s and 1990s, a particular approach called multinomial processing tree (MPT) modeling was developed at a general level by William Batchelder, David Riefer, and Xiangen Hu. The central characteristic of MPT models is that they have a particular type of cognitive architecture represented as a rooted tree structure. Such a structure assumes that cognitive processes follow one another, and subsequent processes are conditionally dependent on the success or failure of earlier processes. For example, if a model has parameters for item attention, item storage, and item retrieval, then successful storage depends on successful attention. In turn, successful retrieval depends on successful storage. If any of these processes fail, then responses may be governed by guessing biases corresponding to various states of incomplete information. Each series of processing possibilities leads to different observable responses, and there are usually many of these processing patterns, each represented by the “branches” of the tree architecture.

**An Example: Batchelder and Riefer’s Pair-Clustering Model**

One early example of an MPT model is the pair-clustering model developed by Batchelder and Riefer. Their model was designed to separately measure storage capacity from retrieval capacity in human memory. The data for the model involve a specially designed free-recall task where participants study a list of words one at a time, and then, at a later time, memory is tested by having the participants recall as many of the studied words as they can in any order. The list consists of pairs of exemplars from several categories such as vehicles (taxi, car) or flowers (rose, daisy). Recall of each category pair is scored into one of four categories: (1) both words recalled successively, $C_1$; (2) both words recalled but not successively, $C_2$; (3) exactly one word recalled, $C_3$; and (4) neither word recalled, $C_4$. The model postulates three parameters each designed to measure a different cognitive process: (i) a storage parameter $c$ representing the probability that a pair of words is “clustered” and stored in memory during study; (ii) a retrieval parameter $r$ representing the probability that a word pair is recalled given that it is stored as a cluster, and (iii) a parameter $u$ for the probability that a word in a pair that was not clustered is recalled as a singleton. Because the parameters refer to probabilities of successful cognitive processes, their values must be between zero and one.

The connection of the parameters to the category probabilities is based on a combination of psychological considerations and reasonable approximations. In particular, it is assumed that both members of a category pair are recalled successively if and only if the words in the pair are clustered and the cluster is retrieved (joint probability $cr$). Also, if a cluster was stored but not retrieved, then neither word is recalled. In contrast, with probability $(1 - c)$, the words in a pair are not clustered, and in this case each word in the pair is or is not recalled individually with probability $u$, subject to the condition that if both non-clustered words are recalled, they are
not recalled successively. These assumptions can be displayed in the processing tree found in Figure 1.

To briefly turn to the mathematical details, it is easy to use this tree to express the category probabilities, $p_j = \Pr(C_j)$ for $j = 1, 2, 3, 4$, in terms of the parameters. The result expresses each category probability as a sum of the probabilities of the branches that lead to that category as follows: $p_1 = cr$, $p_2 = (1 - c)u^2$, $p_3 = (1 - c)u(1 - u) + (1 - c)(1 - u)u$, and $p_4 = c(1 - r) + (1 - c)(1 - u)^2$. It is a matter of simple algebra to show that if we collect data $D = (n_1, n_2, n_3, n_4)$ and estimate the category probabilities by relative frequencies, $P_j = n_j/N$, then we can solve the four model equations for parameter estimates (denoted by *) yielding $u^* = 2P_2/(2P_2 + P_1)$, $c^* = 1 - P_2/(u^*)^2$, and $r^* = P_1/c^*$. In order for these equations to yield estimates of the parameters in the interval $(0,1)$, it is necessary that $(P_3)^2 < 4P_2(1 - P_2 - P_3)$.

Common Aspects of Multinomial Modeling
The example of the pair-clustering model illustrates the basic properties of multinomial modeling, which are the tree architecture and the computational rules that tie the cognitive processing parameters to the categorical data. However, the example does not illustrate three aspects typical of most applications of MPT models. First, in the example, there were three parameters representing cognitive processes and only three degrees of freedom in the data structure (as the four category probabilities are required to sum to one). In cases where there are more degrees of freedom in the categorical data than parameters, the system of equations expressing category probabilities in terms of parameters is overdetermined, and standard techniques in mathematical statistics are used to estimate the parameters. Second, the pair-clustering model involves just one system of categories, but many MPT models are developed for several category systems, each of which is associated with its own processing tree. For example, MPT models for the source monitoring experiment discussed earlier specify three processing trees, one for each item type (Source A, Source B, or New).

Finally, unlike the example of the pair-clustering model, most applications of MPT models involve two or more experimental groups of participants, where the same model with possibly different parameter values is assumed to govern each group’s category count data. In this case, MPT models are used to conduct hypothesis tests in an effort to discover which cognitive processes account for differences between the groups. This approach contrasts with the usual approach in experimental psychology for analyzing data from multiple experimental groups, which is to apply standard statistical tools like analysis of variance or linear regression. Although these tools are well developed to detect group differences and associate them with experimental manipulations, they do not allow one to pinpoint the cognitive bases for the differences.

As can be seen, MPT models are simple statistical models that are easy to develop and analyze. However, before an MPT model can be used as a measurement tool, it must be validated. A validated model is one in which the parameters can be shown to represent the cognitive processes they stand for. Establishing validation involves conducting simple cognitive studies where experimental manipulations are designed to affect some parameters and not others. These experiments attempt to dissociate the parameters by showing that they can be independently manipulated in ways that are consistent with established psychological theory. For example, for the pair-clustering model providing retrieval cues during recall should increase the value of the retrieval parameter $r$ but not the value of the cluster storage parameter $c$. Other manipulations, such as increased study time, should affect the storage parameter but probably not the retrieval parameter.

Current and Future Directions
Since the 1990s, MPT modeling has become an increasingly popular approach to cognitive modeling, and its use has been facilitated by several

![Figure 1](The pair-clustering model)
software packages that can perform parameter estimation and hypotheses testing. To date there have been more than 100 examples of the application of MPT modeling. Most of these applications have been in the standard cognitive areas of memory, reasoning, and perception; however, clinical, social, and developmental psychology are also areas where MPT modeling is active. There are also a number of ongoing projects that explore the statistical properties of these models. For example, there has been recent work creating hierarchical MPT models to handle variation in parameter values due to individual differences in the participants, as well as latent class MPT models that can be used to model subgroups of participants with different cognitive abilities.

William H. Batchelder, Xiangen Hu, and David M. Riefer

See also Memory Recall, Dynamics; Modeling Causal Learning

Further Readings


Multiple Intelligences Theory

Multiple intelligences theory was proposed by psychologist Howard Gardner in 1983 to oppose the narrow meritocracy of g, or general intelligence, and to reflect evidence for the possible brain basis of content-specific information processing. Gardner argued that the general intelligence factor g should be replaced by seven intelligences: linguistic, musical, logical-mathematical, spatial, bodily-kinesthetic, intrapersonal (sense of self), and interpersonal. Gardner proposed that the intelligences were innate and that each intelligence had its own unique brain mechanism comprised of 50 to 100 micro modules.

Gardner’s list and characterization of intelligences changed somewhat over time. In 2000, Gardner proposed an eighth intelligence, naturalist, involving the comprehension of natural things, and suggested that there might be a ninth intelligence, existentialist, involving one’s sense of the relation of the self to the cosmos. In 2004, Gardner suggested two additional mental abilities, which he later characterized as profiles. He suggested that creative specialists in the arts, sciences, and trades have more narrowly focused laser-like intelligence and that leaders with high scholastic IQ may have a more broadly scanning mental searchlight.

Multiple Intelligences Theory Was Applied Widely in Education

Multiple intelligences theory was widely disseminated in the field of education, and many teachers and psychologists published articles that described the application of the theory to classroom practice. Theory adherents claimed that successful classroom application validated the existence of multiple intelligences. Educators and theory supporters argued the varied intelligences provided better cognitive skill profiles of typical students, savants, prodigies, and individuals with brain injuries than did the verbal and visual subskill tests of IQ measures.

Moreover, many educators saw multiple intelligences theory as a means to assess children without using culturally biased standard IQ tests. In
addition, educators valued Gardner’s view that multiple intelligences testing would allow students to be enthusiastic about cognitive assessment. The ideal of multiple intelligences testing was to allow children to solve problems in many different contexts. Standard paper-and-pencil, single-answer tests were thought inappropriate for determining the strengths and weaknesses of the different intelligences, and educators promoted the assessment of multiple intelligences through interesting materials. Teacher observations, student work portfolios, and learning inventories were recommended as a means of assessing the multiple intelligences of a student.

A key corollary of multiple intelligences theory was that assessing all of the intelligences would permit teachers to find each student’s strongest intelligences, and students would have the opportunity to find an intelligence in which they might feel a greater competence. Teachers would then be able to avoid the student’s weaker intelligences and use the student’s stronger intelligences to create learning materials unique to each child. For example, multiple intelligences testing of the naturalist intelligence asked students to play with different natural objects, such as a feather and a stone. Students would be asked to observe the differences and similarities between these objects, describe them in detail, and address why some objects sink and some float. Similarly, musical intelligence was tested by asking students to sing a range of songs from simple to more complex. The assessment recorded whether a student was sensitive to pitch, rhythm, and melody.

**Multiple Intelligences Theory Led to the Meme of Many Types of Intelligence**

The popularity of Gardner’s theory of multiple intelligences led others to theorize individual content-specific intelligences. Six intelligences had currency in different disciplines. Business Intelligence was proposed as skill with processes, trends, tools, and ideas current in business practices. Social Intelligence was hypothesized as four skills: self-awareness, social awareness, positive social beliefs and attitudes, and skill and interest in managing complex social change. Musical Intelligence was characterized as notable sensitivity to pitch, melody, and rhythm. Emotional Intelligence was hypothesized to include perceiving and understanding emotions, using one’s own emotions to facilitate problem solving, limiting negative emotions and maintaining positive emotions. Cultural Intelligence was claimed to include four dimensions: emotional resiliency, perceptual acuity, flexibility and openness, and personal autonomy. Spatial Intelligence was identified as the basis for skill in architecture, because this intelligence was proposed to reflect individual ability to move through varied spaces and recall accurate images of those spaces.

These six theorized intelligences and many additional proposals for unique content intelligences invoked Gardner’s multiple intelligences theory as valid supporting scientific evidence that the brain was indeed partitioned into many possible intelligences.

**Neuroscience Research Has Not Validated Multiple Intelligences Theory**

Gardner argued multiple intelligences theory was validated through its basis in cognitive neuroscience research, through its successful classroom application, and through its ability to successfully account for cognitive skill patterns of individuals. Gardner also claimed that neuroscientists were conducting research to explore the specific brain circuits governing each of the multiple intelligences. It is true that neuroscience research has explored brain circuits and systems underpinning human skills. However, researchers such as Lynn Waterhouse have noted that no neuroscience research had tested the theory of multiple intelligences and that neuroscience research had disconfirmed the existence of the putative separate content processing modules in the brain.

Multiple intelligences theory was never a focus for cognitive neuroscience researchers, because even in 1983 when Gardner initially proposed his model, there was evidence that content processing did not occur in separate brain circuits. For example, in 1983, Robert Dykes outlined the division of components of sensorimotor information and the melding of that information with spatial and visual processing. Nearly 20 years later, Dana Strait, Jane Hornickel, and Nina Kraus confirmed that music, reading, and speech processing occurred in the same shared brain circuits.

Neuroscience evidence from the early 1980s to the present confirmed two major interacting brain pathways for information processing. The ventral path computes the nature of an object, while the dorsal path computes where an object is in space. The music we hear, our spatial analysis of an
architectural plan, our computations to solve a math problem, and our understanding of people's emotions and the content of what they are saying are all processed through these interacting dorsal and ventral brain pathways.

Neuroscience research accumulated significant and compelling evidence for several large information processing systems, each of which may have undermined the possibility that the brain was organized by multiple intelligences. In addition to the dorsal-ventral system, the brain was found to have a fast, automatic decision-making process and a slow, effortful, consciously monitored decision-making process. Jonathan Power and colleagues outlined evidence for a third large brain system that included self-reflection processing of social, emotional, and self-related information, and externally directed information processing active during calculations, listening to others speak, attending an opera, or navigating a kayak. A fourth large information processing system was identified as the action-observation network. It governs automatic imitation and recognition of motor actions, social gestures, facial expressions, and language production of another person. The evidence for all four brain-wide systems demonstrates various pathways of analytic and collective processing of information that Gardner theorized to be separately processed by the individual multiple intelligences.

The Appeal of Multiple Intelligences Theory

Cognitive neuroscience's disconfirmation of Gardner's claims for discrete content processing in the brain had no effect on the dissemination of multiple intelligences theory. It remained popular with the public and continued to be applied in classrooms around the world. In 2009, Ji-Qie Chen, Seana Moran, and Gardner outlined the use of multiple intelligences in classrooms in Argentina, Australia, China, Columbia, Denmark, Ireland, Norway, the Philippines, Romania, Scotland, South Korea, Turkey, and the United States. Chen and colleagues argued that the central benefit of multiple intelligences theory in the classroom was that it allowed a concept to be viewed from multiple perspectives, thus contributing to deeper learning.

Lynn Waterhouse

See also Intelligence, Neural Basis; Learning Styles

Further Readings


**Multitasking and Human Performance**

The pace of modern life places a higher demand on the ability to multitask than at any time in human history. Indeed, today's lifestyle often requires people to manage several concurrent activities and to deal with constant interruptions to ongoing tasks (e.g., ringing phones, e-mail alerts, instant messaging, etc.). Multitasking refers to the concurrent performance of two or more tasks, where each task can be defined as an activity with distinct goals, processes, and representations. For example, it is not uncommon for operators of a motor vehicle to talk on a cell phone or interact with a variety of other electronic devices while driving. In this example, the cognitive operations associated with driving a vehicle (e.g., navigating, maintaining speed and lane position, reacting to both expected and unexpected events, etc.) are independent of the cognitive operations associated with conversing on a cell phone. There is now clear and unequivocal evidence that the current performance of these two activities yields
performance on each task that is inferior to the performance of the two tasks when they are performed separately. In fact, the literature is replete with studies that have examined multitasking performance and found patterns of interference indicating that people cannot perform two or more attention-demanding tasks simultaneously without costs. Why are there limits on human ability to multitask?

Types of Processing
There are two general categories of multitasking activity. On the one hand, a person may alternate performance between tasks, switching attention between them in discrete units of time. This creates a form of serial processing in which the performance of one task creates a bottleneck, briefly locking out the processing of other tasks. There are obvious decrements in performance associated with this processing bottleneck. An alarming real-world example of attention switching comes when a driver attempts to text message on his or her cellular phone. Here, visual attention can be allocated to driving, or texting, but not both at the same time. This may be an extreme example of a bottleneck that involves both cognitive and structural limitations; however, one can find these limitations separately as well. Another example based on a large body of laboratory research using the psychological refractory period (PRP) paradigm has found that the processing of one task systematically delays the processing of another concurrent task. The processing delay is thought to stem from a central-processing bottleneck in information processing that is stubbornly resistant to practice, and empirical efforts to bypass the bottleneck have largely been unsuccessful. Together, these findings indicate that a fundamental characteristic of the cognitive architecture is a limited ability to perform more than one attention-demanding task at a time.

On the other hand, some multitask combinations may allow parallel processing, in which attention is shared between two or more concurrent tasks. In this context, attention has been conceptualized as a resource that can be flexibly shared between concurrently performed tasks. However, given the limited capacity characteristics of attention, there is a reciprocal relationship between the tasks such that as one task prospers, because attention is allocated to its processing, performance on the other task suffers. In 1984, Christopher Wickens developed a multiple-resource model, suggesting that some dual-task combinations may draw on separate processing resources. This multidimensional framework conceptualizes modalities of input (auditory vs. visual), codes of operation (verbal vs. spatial), and modalities of response (vocal vs. manual) as separable pools of attentional resources, with performance in dual-task situations varying as a function of the overlap in demand for processing resources; the less overlap there is, the more proficient the multitask performance will be. Other researchers have suggested that it may be better to consider multitask performance with regard to the degree of cross-talk between the cognitive operations and mental representations of the separate tasks. Interestingly, the cell phone/driving dual-task combination would seem to represent a fly-in-the-ointment for this class of theories because driving is primarily a visual/spatial/manual task and conversing on a phone is primarily an auditory/verbal/vocal task. That is, even though these two tasks should, in theory, draw on distinct processing resources, they nevertheless produce considerable dual-task interference. It seems then that bottlenecks in performance may still be present even in circumstances where parallel processing can, at least in principle, take place.

Task Switching
In addition to the central-processing bottleneck costs, persistent costs are also observed when attention is switched from one task to another. In this case, switching from one task to another results in a cost in reaction time associated with performing the first trial in a series (i.e., the switch cost), and this cost often persists for several trials following the initial switch. That is, it takes time for performance to settle into a stable state following a switch from one task to another. Here again, the costs of switching are largely resistant to the effects of practice. Taken together, a general characteristic of attention switching is that performance, particularly performance requiring speeded reactions, is not as good in multitasking contexts as it is if the tasks are not performed concurrently.

Processing Duration
Distinguishing between switching and sharing models of multitasking often proves difficult based on
empirical data. Indeed, some estimates of rapid switching of attention make virtually indistinguishable predictions from sharing models of attention. Thus, it may be more profitable to think about multitasking in terms of the processing duration of the cognitive operations underlying each task. As the processing duration increases, the evidence for serial processing bottlenecks becomes more compelling (e.g., the text messaging and driving example discussed earlier). Interestingly, one situation where PRP bottlenecks may potentially be bypassed is when the processing demands of the second of two tasks is relatively simple (i.e., with a short processing duration). Moreover, there are suggestions in the literature that practice can improve task efficiency, reducing processing time, and thereby improving multitasking performance. Can practice altogether eliminate the bottleneck in multitasking performance?

**Practice and Automatization**

Under certain circumstances, practice can result in task performance transitioning from slow, controlled, and effortful to fast, automatic, and effortless. In the conditions where performance can be characterized as automatic, researchers have often questioned whether these routines can be performed in combination with another task without measurable cost (i.e., an example of perfect time-sharing). Some of the early research addressing this topic reported that, with high levels of practice, a few hardy souls could learn to play the piano while shadowing words, or to take dictation while reading aloud, although in both instances error rates were higher in dual-task situations than when each task was performed by itself. One of the most carefully controlled studies on this topic claimed that, after extensive practice, subjects were able to perform two visual search tasks without noticeable deficit after one of the search tasks became automatic; however, even in this study, a careful inspection revealed that the accuracy of detection dropped significantly from single- to dual-task conditions. At the other end of the spectrum are situations where there is inherent unpredictability in the environment, such as in the case of talking on a cell phone (where each conversation is unique) and driving (which often requires reacting to unpredictable events). In these cases, performance cannot become automatic even with extensive practice. Indeed, in such cases, persistent costs are observed despite years of real-world practice. Thus, despite some claims in the literature of perfect time-sharing, a critical analysis of these findings suggests that perfect time-sharing, if it exists, is an elusive exception rather than the general rule.

As mentioned earlier, a number of researchers have also explored whether practice can eliminate the bottleneck in performance associated with the PRP effect. In the PRP paradigm, two stimuli from different tasks are presented in rapid succession, and the reaction time to the second of the two stimuli systematically increases as the time between the beginning of the stimuli (the *stimulus onset asynchrony*) decreases. In circumstances where both of the tasks require a manual response, the PRP effects are little changed with practice. However, with separate input and output modalities for the two tasks, participants were able to significantly reduce, but not eliminate, the PRP effects. However, others have argued that, even in the best of circumstances, there is virtually no evidence for perfect time-sharing. Thus, the data from the PRP literature suggest that, even with relatively simple tasks and extensive practice, significant dual-task interference is the rule, at least for the majority of individuals.

**Neural Mechanisms**

There is growing body of neuropsychological evidence that a subregion of the prefrontal cortex, specifically the dorsolateral prefrontal cortex (DLPFC), plays a significant role in maintaining the processing goals for task completion. Switching between tasks places a load on this brain region as one goal, and its related processing operations, is supplanted by another. Moreover, individual differences in the functioning of DLPFC are correlated with differences in multitasking ability. Indeed, a small portion of the population (e.g., 2%–3%) seems to be able to perform complex multitasking operations such as talking on a cell phone while driving with no measurable decrement in performance on either task. Current speculation is that these “supertaskers” use more broadly distributed brain regions to control multitasking performance.

**Threaded Cognition**

Despite the general lack of evidence supporting perfect time-sharing, people do multitask at
performance levels greater than chance (e.g., people do not immediately drive off the road when they talk on a cell phone). What is the mechanism underlying multitasking performance? One group of researchers recently developed an ACT-R based threaded cognition model of multitasking in which the task procedures are entered into a queue and processed in the order in which they enter the queue (i.e., least recently processed rule). Given that productions fire one at a time, threaded cognition creates a central processing bottleneck to the extent that the processing of one production locks out the others in the queue. Moreover, as the processing duration of a thread increases, the apparent costs in multitasking become more apparent. Likewise, as practice strengthens a production (thereby speeding its processing), the lockout period for concurrent operations decreases, reducing the overall interference that this production has on multitasking performance.

**Implications for Multitasking**

It is noteworthy that the distinction between a reduction in dual-task interference and no dual-task interference (i.e., perfect time-sharing) is both practically and theoretically important. The first interpretation suggests that the inherent structure of the cognitive architecture places important, potentially insurmountable, limitations on doing more than one cognitive task at a time. The second interpretation suggests a different structural design such that, with sufficient practice, it may be possible to perform independent cognitive operations simultaneously and without costs. On the whole, the data tend to support the former rather than the latter interpretation for the architecture of mind. Some researchers have speculated that the structural limitations arise from competition for specific brain regions, such as the DLPFC, that are essential for controlling and coordinating cognition and action.

**General Principles**

Several general principles can be distilled from the literature on multitasking. First, with the exception of a few cases, performance in a multitasking context is worse than when each of the tasks is performed alone. That is, performing two or more attention-demanding tasks at the same time always produces interference. Even though many people believe that they are experts in multitasking, a careful analysis of their performance would indicate otherwise. Second, practice can improve multitasking performance in situations where performance on one or more of the tasks can be automated. However, practice does not eliminate all sources of interference associated with multitasking. Third, as the complexity of the constituent processes of each task increases, the impairments in multitasking performance will become more pronounced. In sum, despite the ever-increasing demands placed on attention in the modern world, bottlenecks in the cognitive architecture of the mind/brain place significant limits on the ability to multitask, creating tradeoffs between the quality and quantity of processing.

David L. Strayer, Nathan Medeiros-Ward, and Joel M. Cooper

See also Attention, Resource Models; Attention and Action; Automaticity; Divided Attention and Memory; Psychological Refractory Period

**Further Readings**


Music and the Evolution of Language

The relationship between music and language has inspired discussion and controversy for centuries. Like language, music is a human universal, found in all cultures, and human infants appear to be born with considerable musical abilities. Research on congenital amusia (a severe, heritable form of what is popularly called “tone deafness”) reveals that some humans cannot process the relatively fine pitch differences needed for music perception. A capacity to name notes from memory (absolute or “perfect” pitch) also has a heritable component. These facts suggest that music has some genetic basis. However, unlike language, the function of music remains elusive, and Charles Darwin therefore considered music one of our most mysterious cognitive faculties. Darwin resolved the mystery by positing a hypothetical “musical protolanguage”: an intermediate form of communication that preceded the evolution of language. This entry explores Darwin’s hypothesis and reviews recent data consistent with a close evolutionary link between music and language.

Most contemporary models of language evolution posit some intermediate form of protolanguage. For example, anthropologist Gordon Hewes introduced the term protolanguage in the context of a hypothetical gestural protolanguage, where language initially evolved using the visual/manual modality, and speech evolved later. Another model of protolanguage was offered by linguist Derek Bickerton, who hypothesized a vocal protolanguage involving words and simple meanings, but lacking complex syntax. All protolanguage models share the assumption that language is a complex system whose different components evolved during separate evolutionary stages. Such models differ in the order in which these different components evolved and in the sensory modalities involved.

In Darwin’s model, prehuman hominids were hypothesized to communicate vocally with a system resembling non-lyrical singing. Musical protolanguage was vocal, learned, and emotionally expressive: Vocal utterances could convey emotions and individual identity and would be used during courtship, rivalry, and group bonding. However, this hypothetical protolanguage lacked the kind of explicit propositional meaning that typifies language. It would not be possible to express specific thoughts (like “the red berries on the tree by the lake are poisonous”) in protolanguage, any more than in modern instrumental music. Thus, Darwin’s hypothesis suggests that complex vocal control—proto-speech—evolved before meaning. Note that such vocal performances might include speech-like components (e.g., the “eeny meeny miney mo” of children’s songs, “bebopalula” of rock and doowop, or the meaningless syllables of jazz scat singing): Protolanguage lacked meaningful words but not articulated syllables.

Darwin supported his hypothesis with a number of arguments, many of which are further supported by more recent scientific research. One of Darwin’s main inspirations was song in birds. He recognized that many birds learn their song by listening to other birds sing, a characteristic analogous to humans for both song and speech. This similarity is all the more striking given that most mammals and nonhuman primates do not learn their vocalizations. Apes and monkeys raised in the absence of vocalization nonetheless produce their species-typical calls in the appropriate circumstances. Furthermore, chimpanzees lack the motor control necessary to master song or speech and, even with intensive training, cannot learn to produce even simple monosyllabic words. Thus, the neural mechanisms allowing flexible, learned vocalization in humans have evolved some time since our divergence from chimpanzees.

In contrast, many bird species, including parrots, mynahs, and mockingbirds, can easily imitate speech, environmental sounds, or the songs of other birds. Darwin suggested that vocal learning in humans evolved prior to language and, by analogy, functioned in a communication system similar to that of birdsong. New comparative data unavailable in Darwin’s time support this suggestion: Vocal learning has independently evolved in additional bird species (hummingbirds) and several mammal species (including whales, dolphins, seals, bats, and probably elephants). Although vocal learning allows complex vocalizations to be produced and learned, none of these species use their complex “songs” to communicate propositional information: Vocal learning does not support language, in the human sense. Repeated convergent evolution in many species strongly suggests that a vocal learning capacity can evolve easily, under the right conditions, in
vertebrates, and that it often does so in a “song” context, like that of songbirds or whales.

Further support for the musical protolanguage hypothesis comes from computational characteristics shared by music and language. Both are hierarchically structured: Just as words are made up of separate meaningless syllables, musical phrases are built up from meaningless notes. Just as words can be arranged into longer and more complex sentences, musical phrases are typically combined to form larger melodies. Both systems are capable of generating an unlimited diversity of signals, and only some of these are “grammatical,” obeying unstated rules of combination. Similar properties would have been present in musical protolanguage.

The main difference between music and language concerns the expression of meaning. In language, propositional meanings can be expressed explicitly with great precision. Music, although ill-suited for expressing explicit propositions, is well suited for expressing dynamic, feeling-laden contours, which can be mapped in continuous fashion onto emotions, movement, or other more abstract cognitive dimensions. Thus, music is often better suited to expressing feelings, or eliciting movements like dance, than is language. This aspect of protolanguage has not disappeared in modern humans, and it remains an important function of music. As predicted by Darwin’s hypothesis, spoken language retains a prosodic or “musical” component, and tone of voice, melodic contour, and speech rhythm continue to play an important role in speech (though not the written word) today.

The problem of how propositional meaning came to be associated with protolinguistic utterances has been addressed by linguists Otto Jespersen and (later) Alison Wray with their “holistic” model of protolanguage. In a second evolutionary stage, musical protolanguage was augmented with more specific meanings, but still tied to whole musical performances. Just as “Happy Birthday” signifies birthday celebrations, Jespersen envisioned songs becoming explicit indicators of hunting, gathering, or festivity, or signaling particular individuals or objects. The crucial transformation came when listeners began to detect possible connections between components of propositional meaning and individual subsections of the vocal performance: This was the birth of words as we know them today. Such a process of “analysis” has been demonstrated in computer simulations by Simon Kirby and his colleagues, and can, over many generations, produce a significant vocabulary.

Interest in Darwin’s musical protolanguage hypothesis has undergone a resurgence in recent years; the topic is reviewed by archaeologist Steven Mithen in his book The Singing Neanderthals. New data from neuroscience, musicology, and animal communication have augmented Darwin’s original arguments, and although all models of protolanguage remain controversial, musical protolanguage is a leading model for the evolution of language in contemporary discussions. Further research comparing the neural and genetic mechanisms underlying music and language offers considerable hope in the coming decades for the resolution of these age-old debates about human cognitive evolution.

W. Tecumseh Fitch

See also Language Development; Music Perception; Speech Perception

Further Readings


MUSIC PERCEPTION

Most people assume that the world is just as they perceive it to be. But experiments have forced researchers to confront the reality that this is not the case. What we perceive represents the end of a chain of mental events that give rise to a mental representation of the physical world. Our brain imposes structure and order on certain sequence of sounds and thereby creates what we call music. Molecules in the air vibrate but do not themselves make a sound. Sound—and three of its musical components, pitch, loudness, and timbre—are psychological constructions, present only in the minds of perceivers. Thus, to answer Bishop Berkeley’s famous question, if a
Music Perception

The focus of research in music perception and cognition encompasses the mental and neural operations underlying music listening, music-making, dancing (moving to music), and composing. The science is interdisciplinary, drawing principally on methods from cognitive psychology, neuroscience, and music theory, as well as from musicology, computer science, linguistics, sociology, genetics, and evolutionary biology. Music processing is a complex, high-level cognitive activity, engaging many areas of the brain and employing many distinct cognitive operations. As such, it is a useful tool for understanding functions of the mind and brain and informing larger issues in cognitive science, such as memory, attention, categorization, and emotion.

The field traces its origins to experimentation with musical instruments in ancient Greece and China. Aristoxenus (364–304 BCE) argued that one must study the mind of the listener, not merely the collection of sounds impinging on the ear. In the late 1800s, Hermann von Helmholtz, Gustav Fechner, and Wilhelm Wundt first applied modern scientific methods to study musical experience. Today, music psychology is experiencing a renaissance, with an exponential increase in scholarly activity over the preceding decades. This surge of interest follows increasing communication across scholarly disciplines, the emergence of cognitive psychology in the 1960s, and new technologies that facilitate the preservation, presentation, and manipulation of sound (e.g., magnetic tape, hard disks, computers, digital signal processing).

Building Blocks of Music

Although music can be defined in many ways, most would consider that to be called music, more than one tone must be present, creating a sequence of tones. (We reserve the word note for a tone that is notated, or written on paper.) A sequence of tones spread out over time constitutes a melody; simultaneously sounded tones constitute harmony. Two pitches define a musical interval, and a sequence of intervals define contour—the direction of movement of the sequence (up, down, or same) without regard to the size of the intervals. This kind of directional movement also exists in speech where it is known as prosody. Contour is especially salient and may be subject to preferential processing—infants attend to it more readily than they do intervals, and contour is more easily remembered by adults learning a new melody than are the precise intervals.

Tones typically are written or performed with different durations, and the sequence of durations gives rise to rhythm, tempo (the pace or speed of the piece, loosely related to the temporal interval at which one would tap a foot or snap fingers), and meter (the way in which tones are perceived to be temporally grouped or organized, the most common in Western music being groups of two, three, or four). Our brains assemble these fundamental perceptual attributes into higher level concepts such as melody and harmony (just as a painter arranges lines into shapes, contours and forms in creating a cohesive whole). When we listen to music, we are in fact processing these multiple attributes or “dimensions” and their interactions.

Although our subjective experience of music may be seamless and complete, its perceptual components are processed in separate areas in the brain. For example, pitch height—the dimension of pitch perception that is correlated with frequency—is encoded in primary auditory cortex, which contains a tonotopic map: a map of pitches from low to high that mirrors the neuronal pitch map in the cochlea. In addition to relying on absolute pitch information, human appreciation of music relies on pitch relations; in this regard, human music perception may be qualitatively different from that of most animal species.

Melodies are defined by the pattern or relation of successive pitches across time; most people have little trouble recognizing a melody that has been transposed in pitch. In fact, many melodies do not have a “correct” pitch, but simply float freely in pitch space. “Happy Birthday” is an example of this, typically sung with naive disregard to whether it is being sung in the same key from one occasion to
the next. One way to think about a melody, then, is that it is an auditory object that maintains its identity under certain transformations. So, for example, when we hear a song played louder or slower than we’re accustomed to, we can still identify it.

Music and Evolution

Darwin argued that music was an evolutionary adaptation, used for signaling sexual fitness. The past 10 years have seen an increased interest in questions surrounding the evolutionary origins of music and its relation to language evolution: which came first, or to what extent they may have coevolved. Evidence comes from archeological findings (e.g., bone flutes at ancient burial sites), anthropology (the study of contemporary preliterate and preindustrial societies), biology (especially the study of communication among closely related species, such as chimpanzees), and neuroscience (differential activation of brain circuits by music and language, with music tending to activate phylogenetically older structures). Music is characterized by its antiquity and its ubiquity—no known culture now or anytime in the past lacked music. Music has clearly shaped the course of human culture, societies, and history. In addition, specific neurochemical processes accompany musical activities, including the modulation of dopamine levels in response to music listening.

Musicology and Musical Grammar

Each human culture develops its own traditions for the ways in which its music is created. In general, Western music tends to employ duple or triple meter, corresponding to what we would perceive as a simple march or a waltz, respectively, whereas other cultures routinely employ more complex meter in their music. The system of rules or conventions by which sounds are assembled in a given culture can be thought of as the grammar for that music and as reflecting a musical style, syntax, and idiom. Musical phrases are composed of notes, chords, or both, but as in language, these are not randomly ordered, and a reordering of elements produces a different melody. Some musical experiences that we take for granted are in fact culturally dependent. For example, in Western tonal music, minor chords and keys are often associated with sadness and major chords with happiness, although this is not a cross-cultural universal.

Music and Emotion

Music has been called the “language of emotion,” but it is unclear whether our ancestors used music in this way. Unlike visual art, such as cave paintings, which left a permanent trace for scientists to study, music made an impression only in the minds of those contemporaries who heard it; music recording has existed for scarcely 100 years. Inferences must be drawn from extant writings about music (some of which date to 6,000 years ago) and from the study of contemporary preliterate and preindustrial cultures. To contemporary humans, music represents a dynamic form of emotion—the conveying of emotion is considered to be the essence, if not the purpose, of music and the reason that most people report spending large amounts of time listening to music. Recent laboratory studies have focused on the biological underpinnings of musical emotion, particularly the involvement of neural reward systems. This has been studied through investigating the chill response, a physical sensation up the spine. It varies from individual to individual and is based on a number of factors, such as structural components and loudness of the music as well as character/personality organization of the listener and his or her musical experience. When people listen to music which they report consistently gives them chills, blood flow increases to centers of the brain that are implicated in reward, emotion, and arousal, regions that modulate dopaminergic levels.

Our emotional reactions to music are believed to be caused in part by the meeting and violating of musical expectations. Listeners track the progression of music over time, noting the pitches and rhythms employed, and form subconscious predictions about what will occur next. A musical piece that we find pleasing strikes the balance between meeting those predictions some of the time and violating them in interesting ways the rest of the time.

Musical Preferences

Measures of personality and individual differences have been shown to correlate with taste in music; the correlations are relatively small but significant and robust. From research using the Big Five personality inventory and a cross-section of songs representing major genres and subgenres of Western tonal music, certain consistencies have emerged. Although such research is still in its early stages, results indicate
that extraverted individuals tend to like music that is characterized as energetic and rhythmic. Individuals who rate high on Factor 5, openness to new experience, show no correlation with such music, but rather with music that is described as reflective and complex. Upbeat and conventional music correlates with Factor 2, agreeableness.

Several hypotheses exist as to why musical taste might be related to personality. In some cases, people may prefer and seek out styles of music that reflect and reinforce aspects of their personalities. Personality influences how individuals think, feel, and behave. For example, people with high levels of extraversion may seek out situations that allow them to be talkative and sociable. In contrast, more introverted people tend to seek out environments where they have limited contact with others, especially people they don’t know. Just as people seek out and create social environments that reinforce aspects of their personalities, so, too, might people seek out auditory or musical environments that conform to aspects of their personalities. Because music is a component of social identity in contemporary society, people may also seek out music that they believe will create a desirable impression of them.

Adolescents, in particular, use music as a way of communicating their status and affiliation with a particular peer group or style. Individuals of all ages report using music for mood induction. Those who are normally extraverted, for example, may help to maintain their self-identity and typical level of stimulation by listening to energetic music. Some listen to fast-paced or uplifting music to “get going” in the morning. In times of sadness or sensitivity, individuals may listen to an artist who conveys the same kinds of feelings they are going through, in order to feel understood. Music may also maintain a person’s mood when it is simply music they enjoy.

Daniel J. Levitin and Anna K. Tirovolas

See also Audition, Neural Basis; Emotion Regulation; Music and the Evolution of Language

Further Readings

Naïve Realism

Although the label “naïve realism” has been used for different theories in different contexts, the theories it names tend to be those that play a certain role in a discipline. More specifically, naïve realism tends to be used as a name for a theory that would, within a discipline or subdiscipline, be a starting point—a theory that is motivated by an initial appeal to how things seem to us, pretheoretically speaking. This entry briefly considers naïve realism as it is discussed in cognitive psychology, the philosophy of perception, and metaphysics, and how these different theories are related to one another.

Naïve Realism in Cognitive Psychology

The eminent cognitive psychologist, Ulric Neisser, uses the label naïve realism to name a simplistic view that claims, among other things, that a subject’s visual experience “directly mirrors the stimulus pattern; . . . [is] a passive . . . copy of the stimulus,” and “begins when the pattern is first exposed and terminates when it is turned off” (p. 16). This naïve view is then taken to be refuted by, for example, evidence that expectancy can alter a subject’s visual experience and evidence that the visual experience can persist beyond the extinction of the stimulus.

It is important to note that, although naïve realism is supposed to be a pretheoretic starting point, as Neisser’s characterization shows, this does not stop the views so labeled from being theoretically loaded. On the characterization just given, naïve realism goes along with a piece of cognitive psychological doctrine—that visual experiences represent the world. What distinguishes naïve realism from more sophisticated theories within cognitive psychology is the further claim that the representation is both perfectly faithful and passively created.

Naïve Realism in the Philosophy of Perception

This feature of Neisser’s characterization of naïve realism actually serves to render it inconsistent with philosophical naïve realist theories of experience. Within the philosophy of perception, naïve realism has been used to name two closely related theses. The first is the view that the things that we perceive are actually constituents of the perceptual experiences we have. So, for example, if I see a tugboat, then the tugboat itself—part of the external world—is literally a constituent of my experience. If I merely seem to see a tugboat—say, I have a tugboat dream or tugboat hallucination that is completely indistinguishable from my experience of actually seeing one—because there is no suitably situated tugboat in the world, the experience I have when I merely seem to see a tugboat could not be the same kind of experience I have when I actually see one. The experience of actually seeing something and merely seeming to see something are experiences of fundamentally different kinds. The second claim differs in virtue of being specifically about consciousness: It says that, according to naïve realism, the things that we perceive actually shape the conscious character of our experiences. Thus, according to such a view,
the pink color that we are consciously presented with when we see a flamingo just is the flamingo’s pinkness. Again, this has the consequence that we could not have such a conscious experience of pinkness without there being some pink thing that we are aware of. Both of these claims are inconsistent with naïve realism as Neisser conceives of it because they deny that visual experiences are representations of the world, claiming instead that they are states or events that actually involve the world in some way.

The major objection to these characterizations of naïve realism turns on the possibility of misleading experiences, such as hallucination or illusion. Consider a case in which a subject sees a tugboat. According to this kind of naïve realist, the tugboat is either a constituent of the subject’s visual experience or shapes the contours of the subject’s (conscious) visual experience. Yet, the objection goes, Couldn’t we have an experience just like this in the case of a suitably convincing hallucination? Because hallucinations just are visual experiences that occur in the absence of appropriate objects of perception, a real-world tugboat could not be a constituent of such an experience nor could it shape the conscious contours of that experience. Yet if this experience is just like the experience of actually seeing a tugboat, then doesn’t it show that naïve realism is false: that we can have the kind of experience we have when seeing a tugboat without a tugboat being a constituent of and/or shaping the conscious character of that experience? Defenders of naïve realism therefore need to offer an account of how naïve realism can be made consistent with the possibility of such hallucinations. Similar problems arise when considering the evidence, mentioned above, that the visual experience of a stimulus can persist for a short time after that stimulus is extinguished.

Naïve Realism in Metaphysics

In addition to these theories about visual experience, there is also a metaphysical thesis that does business under the name of naïve realism. This is the position according to which there is a world of physical objects whose existence does not depend on being perceived and that really do possess all the properties—including properties such as color, temperature, texture, taste, and smell—that we perceive them to have (as well as more besides). Although this is a distinct thesis, it is related to the previous two characterizations in as much as, if such claims were true, then metaphysical naïve realism would have to be true too. Given this, if metaphysical naïve realism were shown to be false, then this would show naïve realism about perceptual experience to be false in turn. And over the years, metaphysical naïve realism has been subject to an important criticism. For instance, note that, while orange juice normally tastes quite sweet, the same juice can actually taste quite bitter if it is preceded by something sugary, such as pancakes with maple syrup, say. This suggests that the property of being sweet is not a property that the orange juice has independently of us but is rather a way we experience the orange juice. This kind of contention can also be supported by scientific findings. For example, it has been claimed that empirical research into the workings of our visual systems shows that colors are, so to speak, in the eye of the beholder rather than on the surfaces of objects. Again, then, defenders of naïve realism will also need to show how the scientific world view can be reconciled with a conception of the world as containing such things as colors, textures, smells, and tastes.

William Fish

See also Conscious Thinking; Disjunctive Theory of Perception; Realism and Instrumentalism; Smell, Philosophical Perspectives; Theory of Appearing

Further Readings


Narcissistic Personality Disorder

Narcissistic personality disorder (NPD) is a personality disorder that includes grandiosity, a lack of empathy for others, and a range of behaviors that serve to maintain and increase self-esteem. According to the Diagnostic and Statistical Manual of Mental Disorders (4th edition, text revision; DSM-IV-TR) of the American Psychiatric Association, to qualify as a
personality disorder, the narcissistic emotions, cognitions, and behaviors need to be pervasive (i.e., evident across all aspects of a person’s life) and enduring (i.e., part of a person’s disposition for an extended period of time, beginning in early adulthood). For a person to be diagnosed with NPD, the individual’s personality must “deviate from cultural expectations” in at least two of the following four areas: thought processes (cognition), emotional life (affect), relationships (interpersonal functioning), or self-control (impulsivity). NPD can be related to difficulties in each of these areas (e.g., cognition: overconfidence and resistance to accurate feedback; affect: anger, depression, anxiety; interpersonal functioning: infidelity, aggression, difficulties with commitment; impulsivity: problems related to sensation and novelty seeking such as abuse of illegal substances and gambling). A diagnosis of NPD also requires evidence that the personality traits are causing the individual distress or resulting in impairment. Finally, for an NPD diagnosis to be made, other problems and disorders must be ruled out. For example, the abuse of some drugs such as cocaine may lead to grandiosity and lack of empathy. Likewise, manic states can include grandiosity and overconfidence that may be mistaken for NPD. Diagnoses of NPD should be made by a clinical psychologist, psychiatrist, or trained professional. This diagnosis will often entail a structured interview that covers the specific diagnostic criteria for NPD. The use of clinical interview and inclusion of informant reports (i.e., significant others) may be particularly important with NPD because there is evidence that individuals with NPD lack insight into the nature of their traits. This entry describes the features and prevalence of NPD, its relationship with other personality disorders, treatment issues, and some ongoing controversies.

Specific Diagnostic Features and Prevalence

The DSM-IV-TR describes nine specific criteria for the diagnosis of NPD. To warrant a diagnosis of NPD, a person must meet five of these nine criteria. These (paraphrased) include

1. grandiose self importance,
2. a preoccupation with grandiose fantasies,
3. a sense of “specialness,”
4. a need for admiration,
5. psychological entitlement,

6. willingness to exploit or take advantage of other people,
7. low levels of empathy for others,
8. envy for others or sense that others envy him or her, and
9. arrogance.

Because a full five of nine of these criteria are needed to diagnose NPD, the prevalence of clinically significant NPD is thought to be somewhat rare. The “point prevalence,” or number of people in the general population estimated to currently have NPD in the population, is around 1%. The “lifetime prevalence,” or number of people who have ever had NPD, is higher at around 6%. Some have posited that narcissism and NPD are increasing; recent data suggest that individuals in their 20s have a lifetime prevalence of approximately 9%.

Relationship With Other Personality Disorders

Narcissistic personality disorder is located in the Cluster B group of personality disorders (i.e., NPD, borderline personality disorder, antisocial personality disorder, and histrionic personality disorder). NPD shares characteristics with each of these disorders but differs in important ways as well. For instance, relative to narcissistic individuals, borderline individuals are more anxious, depressive, and impulsive. In contrast, antisocial individuals are more likely to be aggressive and impulsive and often manifest cognitive deficits. Finally, histrionic individuals are more likely to be dramatic, shallowly emotional, and inappropriately seductive.

Narcissism is also related to psychopathic personality. Specifically, narcissism is thought to be strongly linked to the interpersonal and affective characteristics of psychopathy but is less strongly related to the socially deviant behaviors and impulsivity-related traits of psychopathy. Overall, the lines between these personality disorders are not bright. They share many features in common, and often an individual will have characteristics consistent with more than one personality disorder.

Treatment

Treatment for NPD is universally described as difficult. Most individuals with NPD avoid treatment because they experience only limited distress and
tend to externalize blame for their functional impairment. In addition, much of the suffering caused by NPD is actually experienced by individuals who are in close contact with the narcissistic individual (i.e., romantic partners, parents, children, colleagues).

Even when individuals with NPD enter psychological or psychiatric treatment, the dropout rate is high. Individuals with NPD do not react well to criticism, and feedback given in therapy can lead the individual with NPD to end treatment. In addition, practitioners find it difficult to treat individuals with NPD because of problems in establishing clinical rapport.

There is limited empirical evidence as to which treatments work best for NPD. There are reports that various forms of treatment, including cognitive behavioral, interpersonal, schema based, and psychodynamic, can work in some instances. There is no well-established pharmacological treatment for NPD.

**Controversies**

The scientific literature on NPD is relatively small, and there are still many controversies and unanswered questions. One prominent question is whether there are multiple forms or types of NPD. The emerging consensus is that there are at least two forms of NPD. The first is a grandiose form that includes a sense of personal dominance, a callous interpersonal style, and the experience of more positively valenced affect and self-esteem. The second is a vulnerable or covert form that includes more paranoia and envy, a similarly callous interpersonal style, and greater negative affect (i.e., depression and anxiety) and low self-esteem. The DSM-IV diagnostic criteria for NPD appear to focus entirely or predominantly on the grandiose form, although the descriptive text does reference the vulnerable form. A related issue is whether narcissism is a defense against hidden feelings of low self-worth. The data do not support this view. Individuals with grandiose narcissism do not manifest evidence of hidden low self-esteem and individuals with vulnerable narcissism, who do seem more defensive, are aware of their lower self-esteem.

Finally, the etiology of NPD is unclear. Given the evidence supporting the heritability of general and pathological personality traits, it is safe to assume there are some genetic roots. The specific environmental influences are less well understood. For grandiose narcissism, there may be a small link with permissive parenting and noncontingent praise; in contrast, for vulnerable narcissism, there seems to be an association with cold and controlling parenting and some forms of childhood abuse. More research is needed on NPD and narcissism as there has been an imbalance between theoretical speculation and empiricism.

*William Keith Campbell and Joshua David Miller*

**See also** Borderline Personality Disorder; Emotion and Psychopathology

**Further Readings**


---

**NATURAL ACTION SELECTION, MODELING**

Put simply, action selection is the task of deciding what to do next. As a general problem facing all autonomous entities—whether animals or artificial agents—action selection exercises both the sciences concerned with understanding the biological bases of behavior (e.g., ethology, neurobiology, psychology)
and those concerned with building artifacts (e.g., artificial intelligence, artificial life, and robotics). The problem has two parts: What constitutes an action, and how are actions selected?

Models of natural action selection allow us to test the coherence of proposed social and biological theories. Although models cannot generate data about nature, they can generate data about theories. Complex theories can therefore be tested by comparing the outcome of simulation models against other theories in their ability to account for data drawn from nature. Each model attempts to account for transitions among different behavioral options. A wide range of modeling methodologies is currently in use. Formal mathematical models have been complemented with larger scale simulations that allow the investigation of systems for which analytical solutions are intractable or unknown. These include models of artificial animals (simulated agents or robots) embedded in simulated worlds, as well as models of underlying neural control systems (computational neuroscience and connectionist approaches). A potential pitfall of more detailed models is that they may trade biological fidelity for comprehensibility.

General challenges facing models of action selection include the following: Is the model sufficiently constrained by biological data that it captures interesting properties of the target natural system? Do manipulations of the model result in similar outcomes to those seen in nature? Does the model make predictions? Is there a simpler model that accounts for the data equally well? Or is the model too abstract? Are its connections to data trivial, making it too obvious to be useful?

Models of natural action selection have delivered new insights in many domains. What follows is a review of several: (a) the relationship between evolved behavior and optimality, (b) biological mechanisms of action selection, (c) whether or not sequencing behavior can require special representations, (d) the role of perception, (e) explanations of disability or disease, and (f) finally individual action selection in a social context.

**Action Selection and Optimality**

When an animal does one thing rather than another, it is natural to ask *why?* A common explanation is that the action is optimal with respect to some goal. Assessing behavior from a normative perspective has particular value when observations deviate from predictions, because we are forced to consider the origin of the apparently suboptimal behavior. One approach is via the notion of *ecological rationality*: Cognitive mechanisms fit the demands of particular ecological niches and may deliver predictably suboptimal behavior when operating outside these niches. Models assist this approach by determining the behavioral consequences of hypothesized optimal mechanisms. Modelers can also use automated optimization techniques such as genetic algorithms (a machine learning technique inspired by Darwinian selection) to find mechanisms delivering near optimal behavior in specific contexts.

**Neural Substrates**

An important open question is whether there are specialized mechanisms for action selection in brains. Arguably, such a mechanism should have properties including (a) inputs that signal internal and external cues relevant to decision making, (b) some calculation of urgency or salience appropriate to each available action, (c) mechanisms enabling resolution of conflicts between competing actions based on their relative salience, and (d) outputs that allow the expression of winning actions while disallowing losers. Recent computational modeling has focused attention on the *basal ganglia* (a group of functionally related structures in the vertebrate midbrain and forebrain) as meeting these criteria. Other large-scale models encompass both cortical and subcortical mechanisms, indicating that in animals there may be a range of selection mechanisms interacting at different levels of the neuraxis.

**Behavioral Sequencing**

Adaptive action selection requires generating behavioral sequences appropriate to achieve longer term outcomes. Such sequences often appear to have a hierarchical decomposition, with selection taking place at multiple levels of abstraction—from choosing among high-level objectives (e.g., whether to eat, drink, or rest) through to selecting specific movements implementing the same immediate goal (e.g., which grasp to use in picking up a cup). Computational models have explored not only this approach but also the alternative—that apparently...
hierarchical behavior may be implemented by a framework without a hierarchical decomposition.

**Perceptual Selection in Decision Making**

Action selection is mediated by perception as much as by motor control. For example, selective attention can guide action by linking specific motor outputs to one among a range of stimuli. Recent models such as the leaky competing accumulator show that noisy sensory evidence supporting each of a range of alternatives can be accumulated until one option passes a threshold, triggering an action. This model explains experimental data and is mathematically optimal in some conditions. More generally, action selection is sometimes modeled via competing, nested, sensorimotor loops with no clear decomposition into sensory or motor components.

**Disorders of Action Selection**

The normal flow of integrated behavior can become disrupted following neurological damage or disease. Models have suggested that conditions including Parkinson’s disease, schizophrenia, Huntington’s disease, and obsessive-compulsive disorder can be linked to the same corticobasal ganglia circuits that have been identified as possible substrates for action selection. Computational models of these substrates have been used to provide improved explanations for how these disorders arise and to investigate possible avenues for treatment.

**Action Selection in Social Contexts**

In nature, action selection usually involves a social context. Agent-based models of social action selection explore interactions among individuals mediated both directly and indirectly via, for example, resource consumption. Examples include minimalist models of factors that influence the troop structure of primate species and models of how ants determine when and where to move a colony to a new nest; models can even explore patterns of voting in a democratic society. Modeling also allows examination of evolutionary mechanisms operating on individuals that lead to social outcomes.

**Summary**

The study of action selection integrates a broad range of topics including, but not limited to, neuroscience, psychology, ecology, ethology, and even political science. These domains have in common a complexity that benefits from advanced modeling techniques, exemplifying the notion of “understanding by building.” These techniques can help answer many important questions such as why animals, including humans, sometimes act irrationally; how damage to neural selection substrates can lead to debilitating neurological disorders; and how action selection by individuals impacts on the organization of societies.

*Anil Seth and Joanna J. Bryson*

See also Decision Making, Neural Underpinnings; Decision Making and Reward, Computational Perspectives; Layered Control Architectures

**Further Readings**


**Natural Language Generation**

Natural language generation (NLG) systems are computer software systems that automatically generate texts in a human language (English, French, Chinese, etc.) from nonlinguistic input data, using techniques from computational linguistics and artificial intelligence. This entry gives a brief overview of NLG from the perspective of the choices that NLG systems must make, using examples from the specific NLG task of generating weather forecasts.

A number of NLG systems have been built that automatically generate textual weather forecasts. These systems take as input a set of numbers that predict temperature, precipitation, wind speed, and other meteorological parameters at different
locations at various time periods. These numbers are usually produced by a supercomputer that is running a numerical weather simulation model. From this input, the NLG system produces a textual weather forecast that is targeted to the needs of a particular user group; for example, an Arabic text that summarizes marine weather in the Persian Gulf for offshore oil rig workers or an English text that summarizes road icing conditions for local government staff who must decide whether salt and grit should be put on roads.

NLG can largely be regarded as a process of making choices. There are usually thousands if not millions of possible texts that could be produced from a particular data set. For example, consider the temperature prediction data set in Table 1. A number of texts could be generated from this data, including the following:

1. “Tomorrow will be a cool day.”
2. “Temperatures increasing from 10 at midnight to 16 at noon, then falling back to 12 at the end the period.”
3. “Chilly nighttime temperatures will rise to a comfortable 15 degrees by morning. Temperatures will stay at about this level throughout the day, before falling to 12 degrees in the evening.”

Deciding which text to generate requires making several kinds of decisions:

- **Content:** What information should be presented in the text? In the above example, for instance, should the system simply give an overall summary (as in Example 1), or describe how temperature changes throughout the day? In general, an NLG system can communicate only part of the information available to it. Indeed, in some weather applications, the NLG system is expected to generate a few sentences of text from 30 megabytes (!) of input data. The decision on what to communicate is usually based on a model of what is important and significant to the user and often uses artificial intelligence reasoning techniques such as knowledge-based systems.

- **Structure:** How should the information be structured? In Example 3 above, for instance, should this information be communicated in two sentences (as in the texts shown), or should a single sentence be used? More generally, what order should information be presented in and how should information be grouped into sentences, paragraphs, and other document structures? In theory this can be based on linguistic models of document structure; for example, if the NLG system is producing a story, then it should use a narrative structure (e.g., generally order events by the time they occurred at). But in practice, structure is often determined by the genre; this is usually the case for weather forecasts, for example.

- **Lexical and syntactic choice:** Which words should be used to communicate domain concepts? For example, if temperature is going up, should this be lexicalized as *rising* or *increasing*? Similarly, what syntactic structures should be used; for example, should we use simple active voice sentences (as in Example 1), gerund-based sentences (as in Example 2), or some other structure, such as passive voice? In theory, it would be nice to make many of these choices on the basis of psycholinguistic models of reading comprehension, but unfortunately current psycholinguistic models are often not detailed and robust enough to support this.

- **Reference:** How should domain objects and entities be referred to? For example, should we refer to the time 2100 as *end of period* or *evening*? If referring to a previously mentioned object, should we use a pronoun or a definite noun phrase? NLG researchers have developed a number of algorithms for making reference choices, and these appear to work well in many cases, but many areas are under-explored. For example, we have reasonably good models for deciding when to use pronouns but much less satisfactory models for deciding how to refer to times.

<table>
<thead>
<tr>
<th>Time</th>
<th>Temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0000</td>
<td>10</td>
</tr>
<tr>
<td>0300</td>
<td>11</td>
</tr>
<tr>
<td>0600</td>
<td>12</td>
</tr>
<tr>
<td>0900</td>
<td>15</td>
</tr>
<tr>
<td>1200</td>
<td>16</td>
</tr>
<tr>
<td>1500</td>
<td>15</td>
</tr>
<tr>
<td>1800</td>
<td>14</td>
</tr>
<tr>
<td>2100</td>
<td>12</td>
</tr>
</tbody>
</table>
The above decisions can be made in a number of ways. The most common approach is to manually write decision rules, usually based on a combination of (psycho)linguistic theory, corpus analysis of human-authored texts, and discussions (or more formal knowledge elicitation sessions) with domain experts. However, recently, there has been growing interest in trying to learn decision rules automatically, by using machine learning techniques that automatically analyze large corpora (collections) of human-written texts and attempt to infer from the texts the decision rules used by the human writers who created these texts. Such machine learning approaches have been very successful in other areas of natural language processing, such as machine translation.

A general issue in NLG decision making is whether the system should try to make decisions that imitate what human writers and speakers do or whether it should try to make decisions that lead to texts that are optimal in some sense for human hearers. These strategies lead to different systems, since human speakers of course do not generally produce texts that are optimal for human listeners.

The best current NLG systems can produce short texts (such as weather forecasts) that are regarded by readers as being as good as (or even better than) human-written texts. However, no current NLG system can produce long texts (more than one page) that are as good as human texts. This reflects the fact that our current understanding of low-level syntactic, lexical, and reference choices is much better than our understanding of higher level content and structure choices.

Ehud Reiter

See also Conversation and Dialogue; Discourse Processing, Models of; Multimodal Conversational Systems; Production of Language

Further Readings

ACL Special Interest Group on Natural Language Generation: www.siggen.org


Neural Correlates of Consciousness

It is widely agreed that some neural activity must correlate very closely with the occurrence of consciousness and, further, that neuroscience has the ability to discover these neural correlates of consciousness (NCC). This agreement contrasts with the widespread disagreement about whether neuroscience can ever explain how consciousness arises from brain activity.

After defining the NCC, this entry summarizes the two basic approaches to the NCC search, using examples of experimental paradigms. It looks at recent findings and developments and touches on methodological and philosophical controversies.

Definition

Finding the NCC means finding the sufficient neural conditions for the occurrence of consciousness and then narrowing down this set of conditions to the minimally sufficient conditions: the neural activity that most closely correlates with consciousness. For example, blood supply to the brain is part of the sufficient conditions for consciousness but is not a minimally sufficient condition because it is important for much else besides consciousness. Similarly, certain processes in the cerebellum may influence a given conscious state but may not be closely correlated with it because the state may be able to occur even if the cerebellum processes differ. In contrast, it seems that, normally, activity in the inferior temporal cortex (together with other areas) is needed for some aspects of consciousness, and this therefore seems to belong with the minimally sufficient conditions for consciousness. The NCC is normally not defined in terms of finding the necessary conditions for consciousness since it is generally agreed that, at least in principle, different neural states could be sufficient for the same aspect of consciousness. (For example, it should not be ruled out that different neural systems in a developing juvenile brain and an aging brain could correlate with the same type of conscious state.)
Two Basic Approaches to the NCC

The state-based approach to the NCC focuses on the contrast between individuals who are in an overall unconscious state versus individuals who are in an overall conscious state (e.g., being in a vegetative state or being anesthetized vs. being awake). One method is to use brain imaging (such as functional magnetic resonance imaging) to reveal the neural activity characteristic of a patient in an unconscious, vegetative state as he or she recovers and regains consciousness. Such studies suggest that a certain thalamocortical pattern of activity may be part of the state NCC. It is important, but also difficult, to match these conditions as closely as possible such that the observed neural activity in the patient is most closely correlated with regaining consciousness rather than with regaining other kinds of nonconscious, cognitive abilities. This methodology is further challenged by intriguing and disturbing studies that arguably show that some patients in vegetative states in fact are conscious.

The content-based approach takes individuals who are already in an overall conscious state and focuses on the contrast between specific conscious contents within those individuals (e.g., the conscious experience of a face vs. of a house). It is important, but also difficult, to control conditions such that the neural activity observed when content becomes conscious is not correlated with unconscious content processing of the stimulus. For this reason, bistable perception (e.g., binocular rivalry, in which different stimuli are shown to each eye and conscious perception alternates between them rather than blending them) is widely used in experimental paradigms. This paradigm keeps the physical stimuli constant while varying the conscious contents, such that neural changes should be closely correlated with changes in conscious content; other paradigms include blindsight, inattentional blindness, and masking. In bistable perception, imaging and single-cell studies indicate the importance of activity in the inferior temporal cortex but also in many other areas of the brain, including early visual cortex. This distribution of activity makes interpretation of data difficult since early visual cortex is also active during unconscious content processing (unconscious, masked stimuli are processed and can influence behavior, for example). This suggests that qualitatively different types of activity in these same cortical areas correlate with conscious and with unconscious contents; neural decoding techniques are being developed that may assist in distinguishing them.

Beyond Mere Correlates

It is clear that the state-based and content-based approaches to the NCC must complement each other, but it is an open question how. Possibly, advances in our understanding of neural interconnectivity and of basic neurocomputational principles (e.g., appealing to the notion of information integration) will help with this question. This kind of development, and others (such as the emerging neural decoding methods), can be seen as attempts at finding systematic rather than merely “brute” NCCs. This could in the future provide hints to the functions of consciousness itself. Though this may not in the end provide a solution to the problem of how consciousness as such arises from brain activity, it may significantly improve our understanding of the nature of consciousness and provide insights into its various disorders.

Philosophical and Methodological Issues

In most NCC studies, neural activity is picked out indirectly via, for example, fMRI’s ability to pick up the blood-oxygen-level-dependent signal, the significance of which is subject to intense technical debate. Likewise, consciousness is picked out indirectly via behavioral responses such as introspective reports. The reliance on introspection is contentious because there is no independent method for verifying such data. Some believe this disqualifies consciousness, and hence the NCC, as an area of proper scientific research, while others believe that we can have sufficient trust in subjective reports. Thus, whereas the NCC search can be pursued in the absence of a solution to the mind-body problem, it is not entirely innocent of traditional philosophical problems at the heart of the very notion of consciousness, concerning its subjective nature and the absence of direct access to other individuals’ consciousness.

Introspection and the general accessibility of conscious content throughout the cognitive consumer systems (such as introspection, episodic memory, and reasoning systems) in the brain may constitute one notion of consciousness, called access consciousness. Another notion may concern the purely subjective, or experiential, aspect of consciousness called phenomenal consciousness. Some argue that a clear distinction between these notions is needed such that, for example, introspective awareness can come apart from conscious experience; this implies the controversial claim that in principle there can be phenomenally conscious states that the person...
having the experience is introspectively unaware of. In that case, neuroscience should expect to find distinct NCCs for access consciousness and for phenomenal consciousness. Experimental paradigms that clearly distinguish these NCCs are yet to be fully developed, and it is unclear how this distinction relates to the distinction between the state-based and the content-based approaches to the NCC search.

Jakob Hohwy

See also Access Consciousness; Anesthesia and Awareness; blindsight; Consciousness and the Unconscious; Introspection; Sleep and Dreams; subliminal perception; unconscious perception

Further Readings


Neurodynamics of Visual Search

This entry describes the synaptic, neuronal, and cortical mechanisms underlying visual attention. It further discusses how these mechanisms, based on more general principles of competition and cooperation between neurons, underlie the processes involved in visual search.

The Psychophysics of Visual Attention in Search Tasks

The visual system cannot process simultaneously the immense amount of information conveyed in a complex natural scene. To cope with this problem, attentional mechanisms are needed to select relevant scene information. Evidence for attentional mechanisms in visual processing comes mainly from psychophysical experiments using visual search tasks such as those developed by Anne Treisman. In visual search tasks, subjects examine a display containing randomly positioned items to detect a previously defined target. Items in the display that are different from the target are distracters. The main phenomenology can be understood from the dependence of the measured reaction time on the number of items in the display. There are two main types of search displays: feature search or “pop out” and conjunction or serial search. In a feature search task, the target differs from the distracters on a single feature (e.g., only in its color). In this case, search times are independent of the number of distracters. A typical example of pop-out search is the detection of a red bar within an array of differently tilted green bars. The result can be trivially explained with the activation of only parallel processes, and therefore the unique feature defining the target pops out. In a conjunction search task, the target is defined by a conjunction of features; each distracter shares at least one of those features with the target. Conjunction search experiments show that search time increases linearly with the number of distracters, implying a serial process. An example of conjunction search is the detection of a red vertical bar within a display containing vertical green or tilted red bars as distracters; that is, there is only one item sharing simultaneously (a conjunction of) the two features defining the target, but each distracter shares one feature with the target.

The classical hypothesis accounting for attentional selection in vision is that attention enhances the responses of neurons representing stimuli at a single relevant location in the visual field. This enhancement model is related to Hermann von Helmholtz’s spotlight metaphor for focal attention. In this metaphor, a spotlight of attention illuminates part of the visual field; stimuli in the spotlight are processed in higher detail, while information outside the spotlight is filtered out. In this classical view, an object searched for in a cluttered scene is found by rapidly shifting the spotlight from one object to the next until the target is found. In this view, attention is based on explicit serial mechanisms. Based on these concepts, Anne Treisman proposed the so-called feature integration theory of visual selective attention. This theory is based on numerous psychophysical experiments on visual search and offers an interpretation of the binding problem, which asks how features detected by different cortical feature
Neurodynamics of Visual Search

maps are assembled into objects. The feature integration theory distinguishes two processing stages. In the first stage, called preattentive, processing done in parallel across the visual field extracts primitive features (e.g., colors, orientation) without integrating them. In the second, attentive, stage information from limited parts of the field is integrated.

The Neurophysiology of Attention

Recent neurophysiological studies, performed mainly in the lab of Robert Desimone, have shown that the effects of the enhancement of attention on neuronal responses can be understood in the context of competition among the stimuli in the visual field. This neurophysiologically more realistic mechanism for selective attention, originally proposed by John Duncan and Robert Desimone, is called the biased competition hypothesis. The biased competition hypothesis assumes that populations of neurons encoding different locations and features of the multiple stimuli in the visual field are simultaneously activated and therefore are competing. Attention to a stimulus at a particular location or with a particular feature biases this competition in favor of neurons encoding the location or the features attended. This biased competition leads to an up-modulation of the firing activity of the neuronal populations encoding the attended location or features and to a down-modulation of the activity of the neuronal populations associated with distracting stimuli. The attentional biasing comes through feedback connections from areas outside the visual cortex. In this framework, it is clear that the neurodynamics underlying visual search result from a top-down process biasing the competition between neurons in early visual areas, which selects one stimulus in the visual field. In other words, attention is an emergent property of competitive parallel interactions across the visual field.

Precursors for this type of mechanism include earlier neural network models such as the adaptive resonance model of Steven Grossberg and the interactive activation model of James McClelland.

The Computational Neuroscience of Attention

Let us now see in more detail which kind of cortical architecture is behind visual attention. Figure 1 shows a functional cortical architecture that considers attentional top-down processes interacting with bottom-up processes as proposed by Gustavo Deco and Edmund Rolls. In particular, it schematizes how the dorsal (also called where) visual stream (which reaches the posterior parietal cortex [PP]) and the ventral (what) visual stream (via V4 to the inferior temporal cortex [IT]) interact in early visual cortical areas (such as V1 and V2), accounting for many aspects of visual attention. This type of architecture implements at the local and global brain area level the principle of biased competition described above. In the original model of Deco and Rolls, the system was composed of six modules (V1 [primary visual cortex], V2–V4, IT, PP, ventral prefrontal cortex v46, and dorsal prefrontal cortex d46), reciprocally connected as schematized in the figure. Information from the retina reaches V1 via the lateral geniculate.

The attentional top-down signal biasing intra- and intercortical competition is assumed to come from prefrontal cortex area 46 (modules d46 and v46). In particular, feedback connections from area v46 to the IT module could specify the target object in a visual search task; feedback connections from area d46 to the PP module could generate the bias to a targeted spatial location created by a spatial attentional cue in an object recognition task. Each brain area consists of mutually coupled neuronal populations whose dynamics are described by conductance-based synaptic and spiking neuronal models. The equations describing the detailed neuronal dynamics can be further reduced using mean-field techniques, which in this case replace the temporally averaged discharge rate of a neuron with the instantaneous ensemble average of the activity of the neuronal population. The dynamical evolution of activity within a cortical area can be simulated in the model by integrating the population activity in the area over space and time.

The computation of a visual search works as follows. An external top-down bias from prefrontal area v46 drives the competition in IT in favor of the population encoding the target object. Then, intermodular back projection attentional modulation pathways from IT–V4–V1 enhance the activity of populations in V4 and V1 encoding the component features of the target. Only locations in V1 matching the back-projected target features are up-regulated. The enhanced firing of the neuronal populations encoding the location of the target in V1 leads to increased activity in the spatially mapped forward pathway from V1 to V2–V4 to PP, resulting in
increased firing in the PP module in the location corresponding to the target. Consequently, these cascades of biased competition compute the location of the target and are made explicit by the enhanced firing activity of neuronal populations at the location of the target in the spatially organized PP module. This type of cortical model can successfully simulate the neurophysiology and psychophysics underlying feature and conjunction search tasks. Note that the whole simulation is parallel and therefore does not involve a serial moving spotlight process at all. In this model, a conjunction search with more distractors takes longer because the constraints are then more difficult to satisfy and the dynamics of the coupled set of networks show a longer latency to settle. In other words, apparently serial cognitive tasks may in fact be performed by fully parallel processing neuronal networks with realistic dynamics.

Gustavo Deco

See also Attention, Neuroimaging Studies of; Attention and Action; Visual Search

Further Readings


Neuroeconomics

The study of decision making has occupied researchers in neuroscience, psychology, and economics for centuries. Since the late 1990s a group of scholars have begun to combine theories and methods from these three disciplines, forming the interdisciplinary field of neuroeconomics. What these scientists came to believe is that the highly different approaches of
the natural and social sciences to the study of decision making could benefit from mutual interaction and constraint. In one direction, theories from economics and psychology could serve to organize the ever-increasing amounts of neurophysiological data into a coherent framework. In the opposite direction, neuroscientific data could provide biological constraints for models of choice behaviors in economics and psychology. A series of papers, meetings, and books has been shaping this field for more than a decade, attracting general public interest but also criticism from scholars in the three parenting disciplines.

**Background**

The different strategies employed by economics, psychology, and neuroscience in the study of decision making can be thought of as three different levels of abstraction. At the highest level, economists strive to predict human choice behavior based on rigorous mathematical models. These are typically “as if” models, with as if meaning that they do not attempt to unravel the actual decision-making mechanism at an algorithmic level and are only concerned with the behavioral accuracy of the predictions they make. This approach has dominated economics at least since the 1940s when the neoclassical school began to dominate economic thought. Paul Samuelson and others developed this approach around the specification of a set of assumptions or axioms from which choice behavior could be mathematically derived. Based on a small number of axioms as simple as “if a person prefers oranges to apples she will not also prefer apples to oranges,” John Von Neumann and Oskar Morgenstern developed their model for choice between uncertain outcomes (expected utility), which was extended by Leonard Savage to include subjective estimations of outcome probability (subjective expected utility). Von Neumann and Morgenstern also laid the foundations for game theory, in which decisions are affected by choices made by many players with competing interests. Although these models were meant to describe empirical human behavior, their mathematical construction gave them a more normative nature, in the sense that they described how people should behave more than how they do behave.

At the lowest level of abstraction, neuroscientists aim to reveal the neuronal architecture that underlies the same behavior that economists seek to predict. Traditionally, such research was based on brain lesions in humans and animals and on electrophysiological recordings in anaesthetized animals. Those techniques were very helpful in unraveling the neural circuits for simple sensory and motor functions but were not sufficient for the study of higher cognitive functions. The introduction of electrophysiological recordings in awake behaving animals in the 1960s and of noninvasive imaging methods in humans in the 1980s and 1990s enabled neuroscientists to examine neural activity while humans and other animals were engaged in complex behavior. In turn, those technological advances raised a new challenge: Huge amounts of data could now be easily generated, but these data needed to be sorted out and interpreted in a meaningful way.

The gap between the highest and lowest levels of abstraction is bridged by psychologists, who are interested in the mental states and processes, including thoughts and emotions, that are caused by neural activity and that lead to choice behavior. In the second half of the 20th century, both economics and neuroscience embraced concepts and methods from psychology, two processes that eventually resulted in the birth of neuroeconomics.

On the economics side, in the 1950s economists began to pay attention to cases in which human choices deviate from the normative predictions of the mathematical models. Maurice Allais was the first to describe behavior that violates one of the core axioms of the neoclassical school, a behavior better known as the Allais paradox. Allais was followed by Daniel Ellsberg, who described the now famous Ellsberg paradox that violated yet another (although admittedly less critical) axiom. By the 1970s and 1980s the psychologists Daniel Kahneman and Amos Tversky had widely extended these early findings by demonstrating that robust axiomatic violations were more common than has been previously suspected. In a series of experiments they revealed a wide range of choice patterns that falsified one or more of the axioms of expected utility theory, leading many economists and psychologists to conclude that economic models could benefit from psychological data and insights. This realization gave rise to the discipline of behavioral economics, which lay at the boundary of economics and psychology.
At about the same time that these processes took place in the economic world, neuroscientists began to feel the need to use models of mental processes in designing their experiments and analyzing their data. The ability to record the activity of single neurons in awake behaving animals and to image the activity of populations of neurons in humans allowed more than a simple correlation between neural activation and observed behavior. Using models of cognitive function, neuroscientists could now look for correlates of hidden variables of these models. Studies of this type constituted the new discipline of cognitive neuroscience, which lay at the boundary of neuroscience and psychology.

In the mid-1990s these cognitive neuroscientists and behavioral economists who were interested in choice behavior went a step further: Some neuroscientists began to examine economic models and contemplate their use as a normative theory for neurobiological data. At the same time, a few economists considered a further logical reduction of their models, taking into account mechanistic constraints and algorithmic features of the human nervous system. These steps set the stage for the emergence of the new discipline of neuroeconomics.

From Economics to Neuroscience
The first neuroeconomics paper is probably a 1996 review published by Peter Shizgal and Kent Conover in *Current Directions in Psychological Science*, titled “On the Neural Computation of Utility.” The paper explicitly employed normative economic theory to describe the neurobiology of choice in rats performing intracranial self-stimulation. Following this paper, in 1999, Michael Platt and Paul Glimcher published their paper “Neural Correlates of Decision Variables in Parietal Cortex” in *Nature*. The authors showed in that paper that neurons in monkey parietal cortex encoded both the probability and the magnitude of reward, a finding compatible with the idea that these neurons played an algorithmic role in decision making closely aligned with the basic models of neoclassical economic theory.

In 2001, the neuroeconomics approach was extended to human studies with the publication of two imaging studies that resulted from collaborations between economists, psychologists, and neuroscientists. In the first of these papers, Peter Shizgal joined Daniel Kahneman, Hans Breiter, and others to map the neural responses to expectation and experience of monetary gains and losses. Breiter and colleagues based their experimental design on two principles from Kahneman and Tversky’s prospect theory: that the evaluation of a risky prospect depends on its framing as a gain or a loss and that losses loom larger than gains of equal magnitude. In the second paper, a group of economists, including Kevin McCabe and Vernon Smith, together with MRI specialists, were the first to use game theory in a human neurobiological experiment. Playing either against a human opponent or against a computer, subjects had to decide whether to trust the other player. Results showed that those subjects who tended to trust their opponents had higher neuronal activation in regions of the medial prefrontal cortex while playing against humans compared to playing against computers.

Shortly after the publication of these first papers, Glimcher published a book titled *Decisions, Uncertainty, and the Brain* in which he articulated the need for the use of normative theory in the study of the neurobiology of higher cognitive function. Glimcher proposed that economics could provide such theory and called for neuroscience and economics to join forces in the study of decision making.

In the next few years, a growing number of neurobiological papers on humans and other animals that relied on economic theory in the design and analysis of their experiments have been published. Some examples are papers dealing with the economic value of goods, game theory, the framing effect, loss aversion, intertemporal choice, and ambiguity aversion.

While the contribution of economics to neuroscience is widely recognized in the neuroscience community, many economists maintain that economics does not stand much to gain from neurobiological insights. Nevertheless, several recent papers suggest a potential role for the use of neuroscientific data in developing economic theories.

From Neuroscience to Economics
In 2005, the economists Colin Camerer, George Loewenstein, and Drazen Prelec published a paper in the *Journal of Economic Literature*, in which they made the case for neuroeconomics from the
Neuroeconomics side. Camerer and his colleagues argued that understanding the neural mechanism of decision should provide algorithmic insights that will constrain possible economic theories and may direct future studies in economics. In essence, what they proposed was that instead of settling for “as if” models, economists could now use neural data to look for more mechanistic models.

At about the same time, Michael Kosfeld, Ernst Fehr, and colleagues published a paper in *Nature*, which is probably the first to show the potential role neuroscientific data can have in the shaping of economic theory. In that study, subjects played a trust game similar to the one used by McCabe and colleagues described above. Critically in this study, however, in some of the subjects the brain levels of oxytocin, a neuropeptide that is believed to play a role in social attachment, were increased before they made their decision. Kosfeld and colleagues found that those subjects treated with oxytocin were subsequently more trusting compared to a control group.

The next step was taken 2 years later by Glimcher and colleagues, who showed that neurobiological data could be used to falsify existing economic theories. By having subjects make choices between possible gains of different monetary amounts and different times of receipt the authors explored the neural correlates of intertemporal choice. What they found was that the neural architecture underlying this kind of choice process is not compatible with a prominent economic theory of intertemporal choice.

One step further was taken by the neuroscientists Mauricio Delgado and Elizabeth Phelps in collaboration with the economists Andrew Schotter and Erkut Ozbay. In that study, the authors first used neuroimaging to gain insight about the psychological processes that underlie the tendency of most people to overbid in auctions and then used that insight as a basis for behavioral predictions that they tested and confirmed.

Thus, the first attempts to use neurobiological data in the development of economic models have already been made. However, the degree of influence that studies of this sort will have on economic theory remains to be seen. Moreover, there is a strong debate in the economics community not only about whether neurobiological measurements could affect economic models but also whether they should affect those models. Perhaps the most famous attack on neuroeconomics was made by Faruk Gul and Wolfgang Pesendorfer in their article “The Case for Mindless Economics.” Gul and Pesendorfer argued that the goal of economic theories is to make predictions about behavior and that the actual machinery by which choice is accomplished must remain irrelevant to economists.

**An Emerging Field**

A major role in the formation of neuroeconomics was played by several meetings and conferences that allowed scientists from different disciplines to interact and to define themselves as neuroeconomists. This series of meetings eventually led to the formation of the Society for Neuroeconomics. The society has been holding annual meetings since 2005, featuring the most recent studies in neuroeconomics, with the number of participants from around the world rising every year. In 2009, the society published a volume titled *Neuroeconomics: Decision-Making and the Brain*, which was edited by Glimcher, Camerer, Russell Poldrack, and Fehr and authored by all the central scholars in the field. This edited volume summarizes nearly all the most recent advances in neuroeconomics and serves both as an introduction to the discipline and as a handbook for researchers in the field.

Recognizing neuroeconomics as an academic discipline, many universities around the world have opened specialized centers for neuroeconomics, offering both some graduate-level training in neuroeconomics and support for independent researchers. These scientists, together with scholars in traditional departments for neuroscience, psychology, and economics, continue to investigate the behavior and neurobiology of decision making.

*Ifat Levy and Paul W. Glimcher*

*See also* Allais Paradox; Decision Making, Neural Underpinnings; Decision Making and Reward, Computational Perspectives; Decision Theory, Philosophical Perspectives

**Further Readings**


OBJECT-DEPENDENT THOUGHT

Some thoughts are purely general in the sense that they make no reference to specific individual things. *Dogs are descended from wolves, and there are infinitely many prime numbers* are both general in this sense. The first makes a statement about concrete spatiotemporal objects but none in particular; the latter, one about abstract objects but again none in particular. Most of our everyday thoughts, however, are singular or object-directed thoughts in that they make reference to particular individual objects, be they concrete or abstract. Frege was a mathematician; *You are not supposed to smoke in here; It is hot over there; That lime tree is tall; This yellow after-image is fading; I am leaving now; 3 is a prime number.* These are all singular thoughts because each involves reference to a particular thing or things. As these examples indicate, singular thoughts (beliefs, judgments) are usually expressed by sentences containing proper names (e.g., Frege), indexical expressions (e.g., you, I, and now), demonstrative pronouns (e.g., that lime tree, this yellow after-image, here, there), and numeral names (3). The debate over the nature of singular thoughts has been largely restricted to thought about concretely existing objects available to perception. This entry will discuss the controversial doctrine that singular thoughts are object dependent. The following two sections expound the doctrine and note some of its allegedly paradoxical consequences. The next two sections sketch the central argument in favor of object dependence and some objections to it coming from rival conceptions of singular thought.

Singular Thought as Object Dependent

Some philosophers maintain that the mental contents of singular thoughts are object dependent, meaning by this that the existence and identity of their mental contents depend on the existence and identity of the objects those mental contents are about. For example, consider the thought *that is a lime tree* had by you while looking at a particular tree, where the italicized expression specifies the mental content of your thought. According to the doctrine of object dependence, if, counterfactually, no tree at all had in fact been there to be singled out by you, owing perhaps to a referential illusion or hallucination—call this the “empty possibility”—then there would have been no singular thought content for you to entertain. Consequently, your psychological condition in this situation would be different from what it is in the actual situation. Moreover, if, counterfactually, your thought had instead singled out a qualitatively indistinguishable but numerically different tree—call this the “duplicate possibility”—then the resulting thought would have had a different content from the content it has in the actual situation. Again, your overall psychological state in this duplicate possibility is different from what it actually is.

First-person thoughts expressed with the indexical *I* seem clearly to be object dependent. The thought that you now express with the sentence *I am hot* surely could not exist unless you did. Furthermore,
no one else, not even your identical twin, could have had the very same thought. The thesis that singular thoughts expressed with other indexicals, demonstratives, and proper names are object dependent is, however, highly controversial, because of its allegedly paradoxical consequences.

Allegedly Paradoxical Consequences of Object Dependence

It is a consequence of the object-dependent view that a thinker in an empty possibility could suffer the illusion of having a thought when he was not, because his would-be thought failed to pick out an object. But can we really be mistaken about whether we are having a thought? That the answer to this last question is “no” is the very reason why Bertrand Russell notoriously restricted the possibility of genuine singular thought about particulars to those whose existence we cannot be mistaken about—namely, mental entities, such as after-images and other so-called sense data. A person may be mistaken about whether he or she is actually seeing a tree but not about whether he or she is having a visual experience as if of a tree. Another allegedly problematic consequence arises when we consider the duplicate possibility. In such a case, everything will seem the same to you: The duplicate object (a qualitatively indistinguishable tree, say) does not appear to affect your conscious awareness in any way differently from how the actual object affects it. Many of those opposed to object dependence, such as so-called internalists about mental content, argue that in order for there to be a genuinely mental difference between the two cases, you must be able to detect the difference.

The Central Argument for Object Dependence

A number of different arguments have been advanced in favor of an object-dependent conception of singular thought. Many of these, especially those of Gareth Evans and John McDowell, involve a synthesis of key ideas of Gottlob Frege and Bertrand Russell. Some of these arguments are unconvincing to their opponents because they rely on questionable epistemic principles as premises. For example, some of Evans’s arguments appear to rely on an unacceptably strong reading of Russell’s principle that, roughly speaking, to have a singular thought, one must know which object it is that one is thinking about. Sometimes the debate over object dependence hinges on the role of singular thoughts in action explanation. Here, we shall briefly sketch the strongest argument in favor of object dependence.

The argument has three main premises. The first is a very general claim about the nature of thought content—namely, that it is essentially representational, in that it represents the world as being a certain way. When you have what seems to be a (perceptual demonstrative) singular thought, for example, such as that lime tree is tall, your perception-based thought represents the world as being a certain way—namely, that that lime tree (the very one you are seeing) is tall. In other words, there is a certain condition necessarily associated with the thought—the thought’s truth condition—which is such that, if it is fulfilled, then thought is true, and if it is not fulfilled, then the thought is false. The second premise is specific to the nature of singular thoughts: The truth conditions in question must be genuinely singular. What this means is that when the truth conditions are stated, reference must be made, not just to any object fulfilling certain conditions but to a particular object—namely, the very object of your thought. It will not do, according to the object-dependent theorist, to state the truth conditions for your thought as follows: That lime tree is tall is true if and only if there is a lime tree of such-and-such characteristics and it is tall (see the next section for one way of filling out such-and-such). Rather, the truth conditions must make reference to the very tree you are seeing: That lime tree is tall is true if and only if that lime tree (the very one you are seeing) is tall. The third premise has two parts: (a) in the absence of any object (i.e., in the empty possibility), it is impossible for there to be a singular truth condition; (b) in the presence of a different object (i.e., in the duplicate possibility), the singular truth condition will necessarily be different. From these three premises, the object-dependent theorist infers that singular thoughts are object dependent.

Rival Conceptions

Most parties to the debate accept the first premise. The second premise is challenged by those, such as John Searle, who, influenced by Russell’s views, seek to give nonsingular truth conditions based on quantificationally analyzed definite description concepts. A definite description concept is a concept of
Objects of Memory

Memory is a kind of mental state that has an object. Memories are about things. But what entities are the objects of memories? Are those entities mental? Are they states of affairs in the world? This entry will examine three different approaches to this issue. The approach will be philosophical (or conceptual) rather than psychological (or empirical).

Two preliminary points are necessary. First, what counts as the object of a mental state? The key characteristic of memories here is their capacity to be accurate or correct. We can think of the object of a memory as that object, property, state of affairs, or event whose presence makes the memory correct and whose absence makes it incorrect. Second, we need to draw a distinction between experiential and propositional memories. This distinction is drawn in different ways in philosophy and psychology. If you propositionally remember that there is a computer in my office, then you believe that there is such a computer and you believe it because you acquired that belief some time in the past and it has been preserved until now. By contrast, if you experientially remember that there is a computer in my office, then

Sean Crawford

Further Readings
Objects of Memory

you are in a state wherein my office is presented to you as having contained a computer (you have a sort of “memory image” of the computer) and you are in that state because, sometime in the past, you seemed to perceive the computer. The question that will concern us here is what kinds of entities make our experiential memories correct.

The World as the Object of Memory

At first glance, one would think that the objects of memory are worldly entities. After all, we say things such as, “I remember that you were at the party on Saturday,” or “I remember that Jane’s car is blue.” This way of talking suggests that the objects of my memories are those states of affairs that consist in, respectively, your being at the party on Saturday and Jane’s car being blue. Call this the world-directed view. The difficulty for it is the following. Suppose I perceive Jane’s car as being green, but it is really blue. Days later, I am trying to remember what color it was and I happen to have a memory image of it as being blue. Call this Situation 1. Intuitively enough, we would say that my memory has failed me here. However, the world-directed view commits us to saying that, in Situation 1, I am not misremembering the car.

Sensory Experience as the Object of Memory

One is then inclined to turn to the idea that the objects of memory are one’s own past perceptual experiences. This suggests a picture of memory as being similar to introspection. The idea would be that, in both cases, we are attending to our own mental states. In memory, those mental states are past perceptual experiences, whereas, in introspection, they are current mental states. This view, which we may call the introspective view, accounts for our intuitions about Situation 1. However, imagine now that your memory experience of Jane’s car presents it to you as being green. The car is really blue and your memory image originates in a past perceptual experience of it as being green. Call this Situation 2. The introspective view commits us to saying that your memory experience of Jane’s car as being green is correct. But we would not want to say that. If you misperceived the car as being green in the past, how can you be remembering it correctly now when it appears to you as being green? It is hard to see how a false mental state could have turned into a true mental state just because time has gone by.

Memory Is Its Own Object

According to John Searle’s token-reflexive view, the object of a memory experience is a causal relation that involves world and mind. The object of a memory experience wherein Jane’s car appears to you as having been blue is the following complex event: The car being blue caused a perceptual experience of it as being blue, which in turn caused this very memory experience. This happens neither in Situation 1 nor in Situation 2, so the token-reflexive view accounts for the intuitions that our memories are false in those situations. A concern for this view, however, is that it may build too much into the contents of our memory experiences. What if the car is blue, you have a perceptual experience that presents it as being blue, and it elicits in you a memory experience of it as being blue, but it so happens that your perceptual experience was not caused by the presence of the car? Philosophers disagree on whether these perceptual experiences are true. To the extent that you feel inclined to think that they are, this case is a problem for the token-reflexive view, for it commits us to saying that your memory experience is, in this situation, false.

Conclusion

It seems that we have different notions of what memory is supposed to do, and they pull us in different directions regarding the objects of memory. We think that memory gives us knowledge of the past just like perception gives us knowledge of the present. This pushes us toward the world-directed view. We also think that it should preserve perceptual experiences just like propositional memory preserves beliefs. This pushes us toward the introspective view. Despite the fact that the token-reflexive view seems, at first glance, unnecessarily complicated, it seems that some version of that approach might be the best way to relieve the tension between these seemingly conflicting intuitions about memory.

Jordi Fernández

Author’s Note: This research was funded by two grants from the Spanish Ministry of Science and Technology for projects HUM2007-61108 and HUM2006-09923.
Optic Flow

Optic flow refers to the image motion of the environment projected on the retina during our movement in the world. The term was first coined by James J. Gibson and played a key role in the development of the ecological approach to visual perception, an approach that emphasizes studying human perception in the natural environment rather than in a controlled laboratory setting. Ever since Gibson proposed that the optic flow field contains cues for the perception and control of self-motion, much research in cognitive psychology and neuroscience has investigated what specific cues from optic flow people use for the perception and control of self-motion. The major findings are summarized below.

Perception of Self-Motion

The optic flow field is normally represented by a velocity field with each velocity vector depicting the motion of a reference point in the environment. Any optic flow field is composed of two components, a translational component of radial flow, which is the pattern of flow due to the observer traveling on a straight path with no eye, head, or body rotation (pure translation, Figure 1A), and a rotational component of lamellar flow, which is the pattern of flow due to observer eye, head, or body rotation and/or the observer traveling on a curved path (Figure 1B).

See also Causal Theories of Intentionality; Causal Theories of Memory; Intentionality of Bodily Sensation; Intentionality of Emotion; Memory and Knowledge

Further Readings


Figure 1 Sample velocity fields for movement over a ground plane. Each line represents a velocity vector depicting the motion of a reference point on the ground. (A) Translational component of radial flow produced by observer translation toward the x. (B) Rotational component of lamellar flow produced by eye rotation to the right about a vertical axis. (C) Retinal flow field produced by translating toward the x while fixating o on top of a post.

To illustrate, to steer toward a target, we keep the FOE on the target; to stay in a lane during driving, we keep the FOE at the center of road; and to steer to avoid an obstacle, we make sure the FOE is not on the obstacle. Research by William Warren and others has shown that humans can indeed use the FOE in optic flow to estimate their heading within 1° of visual angle during simulated translation. Note that good heading performance for pure translation may not involve the perception of self-motion, because the task could be performed by locating the FOE in the 2-D velocity field on the screen without any 3-D interpretation of the velocity field.

**Translation and Rotation**

When one is traveling on a curved path or is traveling on a straight path but rotating one’s eyes to track an object off to one side, the retinal flow pattern is not radial any more. The flow field now contains both translational and rotational components, and the lamellar flow generated by the path or eye rotation (Figure 1B) shifts the FOE in the retinal flow field away from the heading direction (Figure 1C). To recover heading in this case, many mathematical models have been proposed that use information such as global flow rate and motion parallax in the flow field to compensate for the rotation, a computation that has been implemented with neurophysiological models of primate extrastriate visual cortex.

To determine whether humans are capable of recovering heading from combined translational and rotational flow, a number of behavioral studies have examined heading perception during translation with simulated eye movements (the display is generated in such way that the retinal image of the display on a stationary eye is the same as if the eye had moved). While some behavioral studies by Martin Banks and others show that observers need extraretinal information (such as oculomotor signals about eye movement) to remove the rotational component in the flow field for accurate heading estimation at a high eye rotation rate, more studies by James Cutting, Leland Stone, Li Li, and others find that observers can estimate their heading within 2° of visual angle by relying on information solely from optic flow, especially when a large field of view and realistic complex 3-D scenes are provided.

**Path Perception**

Apart from heading, an equally important feature of self-motion is one’s future trajectory of traveling (path). The common locomotion control tasks that can be achieved using heading can be similarly accomplished using path. Heading and path coincide when one travels on a straight path but diverge when one follows a curved path of motion; in the latter case, heading is the tangent to the curved path (Figure 2).

While heading can be perceived from a single 2-D retinal velocity field of optic flow, path recovery requires more. The instantaneous velocity field during translation and rotation is associated with one heading direction but is consistent with a continuum of path scenarios ranging from traveling on a straight path with eye, head, or body rotation to a circular path with no eye, head, or body rotation. This path ambiguity problem can only be solved using information beyond a single retinal velocity field such as the acceleration in the translational flow field, dot motion over an extended amount of time, reference objects in the scene, or extraretinal signals. All these cues can be used to determine whether the rotational component in optic flow is due to eye, head, body,

---

**Figure 2** An illustration of the relationship between heading and path for (A) traveling on a straight path and (B) traveling on a curved path

or path rotation. However, up to now very few studies have examined how these cues are used for the perception of path trajectory.

For the relationship between heading and path perception, given that heading is the tangent to the curved path (Figure 2) and observers can infer heading as soon as they perceive path, recent studies from Li’s lab have found that while heading and path perception are two separate processes, path does help heading perception when the display does not contain sufficient optic flow information for accurate heading estimation. Furthermore, accurate perception of path but not heading from optic flow depends on where we are looking, thus supporting the claim that heading is a more reliable cue for the online control of locomotion.

**Neural Basis**

Many species have neural pathways that selectively respond to optic flow patterns. The neurophysiological basis of heading perception includes several cortical areas. Earlier single-neuron studies by Charles Duffy, Robert Wurtz, and others report that neurons in macaque dorsal medial superior temporal cortex (MSTd) selectively respond to radial, lamellar, and spiral patterns of optic flow. More recent functional magnetic resonance imaging (fMRI) studies on macaque and human brains by Frank Bremmer, Andrew Smith, David Burr, and others reveal that the ventral intraparietal area (VIP) is also involved in heading perception as well as a human homologue of primate MST, the MT complex (MT+).

For the cortical areas involved in path perception, recent brain-imaging work on humans by David Field and others reports that the presence of road markers, which clearly defined the path trajectory, activates the superior parietal lobe (SPL) bilaterally in addition to the MT+ area. Presenting observers with distant road markers during heading judgment reproduces the SPL activation, whereas presenting observers with near road markers results in activation only in the MT+ area.

**Control of Self-Motion**

Gibson proposes that we use the information that we perceive from optic flow to guide our movement in the world. The main research findings on optic flow cues used for visual feedback-driven control of self-motion are summarized below.

**Walking Toward a Target**

James Gibson states that to steer toward a target, we move in such a way to keep the FOE in optic flow (i.e., heading) on the target. However, work by Simon Rushton and others has challenged this claim. They find that when observers wearing a prism are asked to walk toward a target, they walk on a curved rather than a straight path. The prism deflects the visual direction of the target from the observer, but it does not deflect the FOE in the optic flow pattern from the target. The results thus support the idea that observers rely on the visual direction of the target but not the FOE in optic flow to walk toward the target. Nevertheless, testing people in a virtual environment in which optic flow information can be rigorously controlled, recent work by William Warren and others finds that both the FOE in optic flow and the visual direction of the target contribute to control of locomotion on foot. The FOE appears to increasingly dominate control as more flow and motion parallax information is added to the scene.

**Braking**

The rate of expansion in optic flow specifies the time-to-contact with objects and can thus be used for the control of braking during driving. David Lee proposes that by adjusting deceleration so that the rate of change in time-to-contact is near the margin value of −0.5, one would stop at the moment of contact. Several naturalistic studies by Lee and others report that hummingbirds indeed follow this strategy in docking on feeding tubes. A behavioral study from William Warren’s lab also confirms that observers adopt this strategy to control the direction and magnitude of braking for a linear brake with no higher order control dynamics during simulated driving. However, recent work by Brett Fajen shows that observers do not rely on a single optical variable for braking control during driving. As the dynamics of the controlled system influences the visual cues observers see in the display due to their control actions, observers rely on different optic flow cues (such as global flow rate) to modulate deceleration during braking depending on the dynamics of the braking system (i.e., the mapping between brake position and deceleration).
Optic Flow

Lane Keeping

Another commonly experienced control of self-motion task is lane keeping on a straight path during driving, riding a bicycle, or walking down a path. There are at least three types of cues from optic flow that we can use for lane keeping. The first one is, again, the FOE in radial optic flow. Lane keeping can be achieved by keeping the FOE (i.e., heading) centered on the far end of the lane. The lane edges provide two other cues for lane keeping: bearing and splay angle. Bearing refers to the direction from the observer to a reference point on the lane edge, measured with respect to a reference direction such as the gaze direction or meridian, and splay angle refers to the angle between the optical projection of the lane edge and a vertical line on the image plane. To maintain traveling in the center of a lane, observers can adopt the strategy of keeping the left and the right bearing or splay angle equal. The further away the reference point on the lane edge, the less useful bearing information because the harder it becomes for the observer to detect a change of bearing. On the other hand, as the near and the far parts of the lane edges provide the same splay angles, unlike bearing, splay angle information is a property of the whole image plane, independent of distance.

The FOE in the radial flow, bearing, and splay angle strategies for lane keeping in the real world are usually redundant and lead to the same lane-keeping behavior. Early research in human factors has reported that human operators use heading more than the vehicle’s lateral position (which defines bearing and splay angle) for lane keeping. Later work by Andrew Beall and Jack Loomis has found that people rely mainly on the splay angle for lane keeping. Recent work by Li Li challenges this finding and shows that heading from optic flow is used for lane keeping regardless of the presence of splay angle information. Several other studies reveal that equating the speed of radial flow in the left and right lateral field of view also contributes to maintaining a centered position in the lane.

In summary, in support of Gibson’s proposal that optic flow contains cues for the perception and control of self-motion, research in cognitive psychology and neuroscience over the last four decades has not only identified the cues in optic flow that we use to perceive and control our self-motion in the world but also the underlying neural mechanisms responsible for the detection of these cues. As our detection of information in optic flow puts us in direct contact with the world without the need of mediating representations, optic flow provides the key supporting evidence for the concept of direct perception.

Li Li

See also Common Coding; Depth Perception; Motor System, Development of

Further Readings


PERCEPTUAL CONSCIOUSNESS AND ATTENTION

The topic of perceptual consciousness has been of great interest to researchers in recent years, ever since the surge of research on what is called implicit perception—that is, perception without awareness. Prior to that, most researchers would have considered the term perceptual consciousness redundant since the term perception was generally defined as the awareness of objects and events through the medium of our senses. Defined in this way, there simply was no perception without awareness. It was only when evidence began to accumulate that we could, in some sense, perceive objects and events without being aware of them—in other words, implicitly perceive them—that serious interest in perceptual consciousness emerged. The primary evidence for implicit perception derives from evidence of what is called priming—that is, that an object or event present in our environment can affect our subsequent behavior. For example, if a picture of a corn stalk is flashed in front of us so quickly that we do not see it, if we now are shown the word stalk and asked to define it, the first definition we are likely to give is “something on which corn grows” rather than “the act of following someone obsessively.” This is so even though the latter is the more frequent use of the word and is the definition most frequently given by observers who have not been primed by the flashed picture of the corn stalk. Evidence of perceptual priming forces us to distinguish between perception that is implicit and perception that is explicit—that is, which entails awareness and which is referred to as perceptual consciousness.

The Relationship Between Perceptual Consciousness and Attention

If perceptual consciousness refers to those perceptions we are aware of, what is the contribution of attention to these kinds of perceptions? One frequently given answer, supported by a very large body of experimental evidence, is that attention is necessary for perceptual awareness. In fact, there are a number of perceptual phenomena that strongly suggest this. One of them is inattentional blindness, a phenomenon first reported by Arien Mack and Irvin Rock, which refers to the failure to consciously perceive an unexpected object that may appear exactly where one’s eyes are focused when attention is engaged in some other task. As shown by Daniel Simons and Christopher Chabris, this can occur even when the unseen event would seem to be highly salient—for example, a man dressed in a gorilla suit walking across a room in which people are passing around a basketball. The gorilla may not be seen when the observers are busy counting the number of passes among the players wearing either the white or the black uniforms. Another closely related phenomenon is that of change blindness, which refers to the frequent failure to perceive a change in a scene that you are viewing and monitoring for changes. The standard technique for demonstrating this phenomenon in a laboratory, devised by Ronald Rensink and his colleagues, involves presenting an alternating pair of real life scenes separated by gray, blank
Perceptual Consciousness and Attention

fields. (The interleaved blank fields serve to eliminate the motion transients that otherwise would signal the presence of a change.) The scenes are identical except that one or more elements are removed or changed in one of the pictures. Change blindness frequently occurs unless the change(s) is central to the gist of the scene. Another closely related phenomenon is the attentional blink, which is normally demonstrated by experiments in which observers are asked to search for two sequentially occurring targets embedded in a series of rapidly presented other items. If the second target appears between 200 and 500 milliseconds of the first, it is very likely not to be perceived.

Sighted Blindness

The above phenomena, which can be considered instances of sighted blindness, that is, blindness that occurs in normally sighted observers looking at above threshold stimuli, are thought to be causally related to the absence of attention. In the case of inattentional blindness, the observer is engaged in some task that requires attention when an unexpected stimulus appears that, because it is not the object of attention, is not seen. Change blindness also is attributed to an absence of attention although it involves more processing steps than inattentional blindness. In change blindness the relevant elements in one array must be encoded in memory and remain there long enough so that they can be compared with the comparable elements in the other array, while inattentional blindness simply involves detecting the presence of a new object. The fact that changes which affect gist are likely to be detected is consistent with the failure-to-attend account of change blindness since the meaning of the scene is what we are most likely to pay attention to. Finally, in the case of the attentional blink, the second target is missed because attention is still engaged in processing the first one and so is unavailable for the processing of the second one, which is therefore not seen. In each of these instances, it is the absence of attention that leads to the failure to see something that is completely visible. It therefore would seem to follow that attention is a necessary condition for perceptual awareness.

Visual Neglect

At least one other phenomenon also lends support to this conclusion, but unlike the three already mentioned, this additional phenomenon only occurs in people who have suffered cortical damage (usually to the right cerebral hemisphere, more specifically to the right posterior parietal lobe). This disorder is called either unilateral, hemispatial, or visual neglect. It is characterized by the failure to see (i.e., to consciously perceive) objects on the left, located opposite the side of the lesion despite the fact that if the objects were located on the right, they would be seen. All explanations of this failure to see point to inattention—that is, the inability to attend to objects on the left. So here again, inattention is invoked to account for the failure to consciously perceive objects that are otherwise visible.

Attentional Load and Perceptual Consciousness

Additional evidence of the centrality of attention for conscious perception comes from evidence, primarily gathered by Nilli Lavie and her colleagues, relating conscious perception to attentional load. This evidence suggests that the greater the attentional load, where load refers to the amount of attentional capacity required by a task, the less likely are stimuli that are irrelevant to the perceptual task to be seen. This again testifies to the importance of attention for conscious perception.

Is Attention Either a Necessary or a Sufficient Condition for Perceptual Consciousness?

There are a series of difficult questions that arise from the linking of conscious perception to attention. If attention is necessary for conscious perception, then it must follow that there are no instances of perceiving that do not entail attention. Furthermore, if it is necessary, then where in the processing stream leading to conscious perception does it operate? Is paying attention also a sufficient condition for seeing? That is, does attending always entail seeing, which, if true, would make visual awareness and attention indistinguishable?

It is difficult if not impossible to find an instance in which conscious perception is independent of attention. One example, suggested by Victor Lamme, comes from the study of binocular rivalry, which arises when the images viewed by the left and right eyes are different. It has been suggested that what we see under these conditions is not a function of attention. There is, however, considerable evidence that attention does influence which of the
two rivalrous images is seen, so this does not seem to qualify as a persuasive example of conscious perceiving without attention. In the absence of such evidence, the hypothesis that attention is necessary for conscious perception stands.

While it is difficult, if not impossible, to come up with instances in which conscious perception occurs in the absence of attention, it is not at all difficult to come up with instances of attending and not seeing anything. For example, paying attention to a location in space where there is nothing to see, or if what there is to see is not visible either because it is presented too briefly or of too low contrast, does not result in a conscious percept. This therefore must mean that while attention is necessary for conscious perception, it is not sufficient for it, although in 2008 Robert Kientrige and his colleagues presented evidence that attending to below threshold or masked stimuli increases their capacity to act as primes. This latter kind of evidence does indicate that attention affects processing even when no conscious percept occurs.

**How Does Attention Enable Conscious Perception?**

Assuming that attention is necessary for conscious perception, at what point does it operate and how? These are large questions but brief descriptions of suggested answers are possible. The first question has led to two competing accounts of where attention operates in the processing of visual input. In one account (early selection theory), attentional selection occurs early in the processing of inputs, while in the other (late selection theory) it occurs late. In a third account (see above), the level at which attention operates is a function of the attentional load, such that the lower the load, the later the filtering and consequently the more input that is consciously perceived. The preponderance of current evidence suggests that the attentional filter occurs late, after meaning has been analyzed. This evidence is seen in findings showing that highly meaningful stimuli, such as one’s own name, are seen even when viewed under conditions of inattention, which normally cause inattentional blindness or the attentional blink. Moreover, Arien Mack and her colleagues have reported evidence that such stimuli are even seen with a high attentional load.

**Attention Amplifies the Input**

The most common answer to the second question about what attention adds that causes a stimulus which is the object of attention to be perceived is that attention amplifies the neuronal responses. In other words, it enhances the level of activation produced by the stimulus, making it more salient. This is thought to be analogous to the way in which increasing stimulus contrast increases stimulus salience. This view is consistent with evidence showing that variations in the direction of attention lead to qualitative differences in performance comparable to those found when stimulus quality is varied. It is also consistent with the proposal made by Petra Stoerig and Alan Cowey in 1997 that the cortical damage that produces visual neglect leads to degradation in the quality of the representation of the contralateral input of the sort generally associated with the absence of attention.

**Does Attention Foster Conscious Perception or Memory?**

If attention acts as an amplifier of input, does this make the representation of the stimulus more salient, which is why we are aware of it, or does attention only enable the encoding of the attended input into working memory so that we are able to report it? The difference here is whether the amplification of the neural response to the stimulus, which is said to be a function of attention, increases the likelihood of our perceiving it or only of our reporting it. If it were the latter, then our failure to report a stimulus to which we are inattentive would not be due to our failure to see it but only to our failure to consciously remember and report it. It would be what Jeremy Wolfe has called inattentional amnesia. If it were the former, then it would be a true failure of perception and rightly termed sighted blindness or inattentional blindness. There is some disagreement over which of these two accounts is correct.

**Priming**

As mentioned at the outset, there is considerable evidence that stimuli we do not see because they are flashed too quickly or because we have not attended to them, as is the case in inattentional blindness, the attentional blink, and unilateral neglect, are capable of priming. It is not clear, however, which account
this kind of evidence supports. What is clear from this evidence, however, is that the unseen stimuli are processed and encoded to some extent by the perceptual system in the absence of attention; otherwise they could not act as primes.

Arien Mack

See also Attention and Consciousness; Attentional Blink Effect; blindsight; Change Blindness; Inattentive Blindness; Psychological Refractory Period; Visual Masking

Further Readings


PERCEPTUAL CONSTANCY

Perception concerns the relationship between physical properties of the world and our conscious experience of them. One area of great interest to perceptual psychologists is perceptual constancy. Perceptual constancy concerns the degree to which a perception remains the same under varying conditions. Research in this area typically involves keeping some characteristic of a stimulus physically constant and asking observers to make judgments concerning that characteristic under varying contextual conditions.

Researchers have studied color constancy as a function of illumination, color surround, chromatic adaptation to prolonged exposure to a color, and off-color objects known to have a specific color (such as an orange-colored cherry). Others have examined lightness constancy as a function of overall illumination, shadows, and brightness of the surround. Another area of interest is shape constancy where researchers have examined perceived shape as a function of object orientation and distance from the observer to the object. Others have looked at slant constancy as a function of object shape and configuration. Space is too limited to fully describe all aspects of perceptual constancy here, so this entry focuses in detail on the oldest and most thoroughly researched area of perceptual constancy research: size constancy.

Size Constancy

In the typical size constancy experiment, a comparison stimulus of adjustable size is located near the observer, and a standard stimulus of constant size is located at several distances from the observer. The observer’s task is to adjust the comparison stimulus until its size matches each standard. As one would expect, observers accurately reproduce the standard when it is close to the observer (at the same distance away as the comparison). However, adjustments can become increasingly erroneous as the standard grows more distant from the observer.

If the observer accurately adjusts the near comparison to match the standard at all distances, the data are said to show constancy. On the other hand, if the observer sees distant standards as being smaller than they really are and adjusts the comparison to be physically smaller than the standard, the data are said to show underconstancy. Conversely, if distant standards dispose the observer to make the comparison too large, the data are said to show overconstancy.

Historical Review

Scholars have been interested in size constancy since ancient times. Euclid (c. 300 BCE), Ptolemy (2nd century CE), Plotinus (c. 300 CE), Ibn al-Haytham (c. 1030), René Descartes (1637), and Leonardo da Vinci (c. 1500) all described their observations concerning size constancy and offered explanations for the phenomenon.
Empirical work on size constancy began in earnest during the late 1920s. Egon Brunswik supervised much of this research and provided the first theoretical account for size constancy. Brunswik believed that size judgments reflect a compromise between an objective attitude, where the observer attempts to adjust the comparison to accurately reflect the physical size of the standard, and a subjective attitude, where the observer defines size as an artist would and attempts to adjust the comparison so that its visual angle matches the visual angle of the standard. Here, the visual angle refers to the percentage or proportion of the field of vision taken up by the standard, not its physical size.

Since all size judgments should reflect some combination of these two attitudes, Brunswik suggested that all size judgments should vary between constancy, when the objective attitude predominates, and strong underconstancy, when the subjective attitude does. Overconstancy is theoretically impossible, and any data that show overconstancy must result from experimental error.

Unfortunately, empirical research of this period often was at odds with this theoretical formulation. First of all, observers seemed incapable of fully assuming the subjective attitude. When asked to assume a subjective attitude, observers adjusted the comparison to be much larger than it should have been to correctly match the visual angle of the standard. Brunswik described this phenomenon as a “regression to the real,” and it showed that observers are incapable of fully assuming a subjective attitude and ignoring physical size. Trained, intelligent, or artistic subjects were better at assuming the subjective attitude, but even these subjects could not fully overcome regression to the real. Second, a number of empirical studies of this era resulted in theoretically impossible results indicating overconstancy.

Early research analyzed each standard separately. Size constancy methodology changed significantly after Alfred Holway and Edwin Boring introduced a new research paradigm in 1941. They asked five observers to judge the apparent size of circular standards located between 10 and 120 feet from the observer by adjusting a variable comparison located 10 feet away under four viewing conditions. Under binocular conditions, they found a slight overconstancy (the comparison was made too large for distant standards), monocular viewing resulted in constancy, monocular viewing through a small window showed strong underconstancy, and monocular viewing through a dark tube displayed very strong underconstancy that almost (but not quite) achieved a projective match, where the visual angle of the comparison would actually be equal to the visual angle of the standard. Subsequent research followed their multiple standard methodology.

Summary of Empirical Research

In 2006, Mark Wagner conducted a meta-analysis of size constancy research (totaling 118 data sets) from the time of Brunswik to the present, the majority from the post-Holway and Boring period. A number of factors were found to influence size constancy data.

For frontally oriented targets under full-cue conditions, objective instructions (which require the observer to adjust the comparison to physically equal the standard) show overconstancy, the most distant standard was seen as being +28% too large. On the other hand, apparent instructions (which ask the observer to adjust the comparison to look or appear equal to the standard) approximate constancy, on average displaying a slight underconstancy of −2%. Projective instructions (which ask observers to match the visual angle of the standard) show marked underconstancy, averaging −37% for the most distant standard.

Reduced conditions (such as viewing in a darkened room or through a small window) are associated with underconstancy for both objective and apparent instruction. When cues to depth are completely eliminated by controlling for the illumination of nearby objects, judgments approach a visual angle match.

Flat stimuli (oriented parallel to the ground) under full-cue conditions also show strong underconstancy, averaging −30%. For projective instructions, this overconstancy averages −70%.

Another factor that appears to influence size judgments is age. Young children usually display underconstancy with an increasing tendency toward overconstancy with increasing age.

Explanations for Size Constancy Phenomena

The standard explanation for most size constancy research is William Epstein’s size-distance invariance hypothesis (SDIH). The great majority of studies support the SDIH and find that perceived size \( s' \) is
related to perceived distance ($d'$) and the visual angle subtended by the object ($\theta$) by the equation

$$s' = d' \tan \theta$$

However, other explanations for size constancy phenomenon have been offered, including V. R. Carlson’s perspective size hypothesis. According to this theory, observers are aware that in some sense objects appear to shrink as they get farther from them observer. Objective instructions incline the observer to overcompensate for this effect, while apparent and projective instructions cause observers to embrace this effect in differing degrees. Suzanne McKee and Harvey Smallman propose the dual calculation hypothesis, which suggests that perceived size not only depends on perceived distance but also on surface texture and the inclusion of objects of known size. Instructions can alter the relative weights given to these different sources of information. John Baird and Mark Wagner are able to successfully mathematically model past size constancy research with their transformation theory. In this theory, the physical sizes of stimuli are transformed into visual angles at the retina. To recover perceived size, the visual system must engage in an inverse transformation. However, this inverse transformation can lead to errors in perceived size if the visual system does not register the correct distance to and orientation of the stimulus. Instructions, cue conditions, stimulus orientation, and age can influence one or the other of these factors.

Mark Wagner

See also Depth Perception; Stereopsis; Visuospatial Reasoning

Further Readings


### Personal Identity

This entry is concerned with the concepts of a person and of personal identity and with various theories of persons and their identity. A paradox is unearthed: We have strong reasons for thinking that any (human) person is identical to a human being yet equally strong reasons to deny this. Our concept of a person thus harbors a contradiction.

**Some Terminology**

The phrase personal identity means different things in philosophy, psychology, and everyday life. Our concern here is with the phrase as it has been understood by philosophers. Typically what philosophers who discuss personal identity want to know is what makes it the case that a person at one time is the same as a person at some later time. Though less frequently discussed, philosophers have also wanted to know under what conditions a single body houses one person or two (as in, e.g., cases of split personality).

Since the phrase personal identity derives from person and identity, these two notions are conceptually prior to that of personal identity. However, this doesn’t mean that they are prior in every sense. Certainly, if we know what a person is, then we know what it is for the same person to persist through time. Still, it may be that the best way to discover the nature of persons is by sifting through competing theories of what it is for a person to persist through time. If we know what changes a person may or may not survive, we will have more idea of what kind of thing is a person. This is what motivates and justifies the methodology of thought experiments.

Some comments on identity and person: In this discussion, we mean identity in the sense of strict numerical identity, not in the sense of qualitative identity. So we are not concerned with the sense of
identical as it appears in identical twins. Twins may be very similar (qualitatively identical), but numerically they are two people, not one. Note also that in talking of identity our concern is metaphysical not epistemic. That is, we are not asking about the kind of evidence we typically rely on in making judgments of personal identity (e.g., physical appearance). Our concern is with what, if anything, constitutes personal identity. What makes it the case that a person at one time is the same (numerically) as a person at some later time? The answer to this constitutive question will not be the same as the answer to the evidential question, since evidence such as appearance, or even fingerprints, is never a logical guarantee of personal identity.

What of the term person? Why do we have such a term? What distinctive work does it do? It would not be too controversial to maintain that we use person to delineate a certain kind of mental being—namely, a self-conscious mental being. This definition puts no restriction on the kind of entity that can be a person. As far as the definition goes, persons could be bodies, brains, nonphysical souls, robots, Martians, parrots, dolphins, or creatures yet to be encountered. This liberality is a strength but also a weakness. For one might have hoped that an answer to the question “What is a person?” would tell us what ontological category (or category of being) we belong to; that is, it would answer the question “What are we?”

**What Is a Person?**

Here then are some ontologically committing answers to the question “What are we?” We are the union of a nonphysical soul and a physical body (René Descartes’s answer); we are animals, specifically human beings (the answer of contemporary Animalists such as Paul Snowdon and Eric Olson); we are not substances (psychic or biological), rather we are “bundles of perceptions” with no substantial self to bind or unify members of the bundle (David Hume’s answer). Hume famously compared the self to a republic, an idea endorsed recently by Derek Parfit and called by him “reductionism about persons.”

Clearly these answers would also answer the question of personal identity. If persons are souls, then what it is for that person to persist is for that soul to persist; if persons are human beings, then what it is for that person to persist is for that human being to persist; if persons are bundles of perceptions, then what it is for that person to persist is for that bundle to persist (though, unlike the persistence of a substance, the persistence of a bundle is a largely conventional matter). In this sense, the question “What are we?” is prior to the question “What is it for the same person to exist through time?” Let’s examine various answers to the first question.

**Dualism**

Many modern philosophers are dissatisfied with Descartes’ answer, for a variety of reasons. First, Descartes’ dualism of soul and body may have seemed reasonable when mentality admitted of no other explanation. But now that we increasingly understand mental activity in terms of brain activity, the need to postulate a nonphysical soul as the bearer of mental states has vanished. Charles Darwin’s discovery, two centuries after Descartes, further reinforced this effect. Second, many philosophers have claimed not to understand how a nonphysical soul and its nonphysical mental states are supposed to interact with the physical realm. For interact they plainly do: Stick a pin in your leg and you cause a sensation of pain. Descartes tried to address this worry, but few find his response plausible.

**Hume’s Answer**

What of the Humean answer? In Derek Parfit’s version of it the key idea is this: Just as the people and land that make up a republic can be understood without reference to the concept of a republic, so a person’s body and mental states (elements of the bundle) can be understood without reference to the concept of a person. (Only so can we think of the self as reducible to the bundle.) However, Parfit’s Humean view faces serious problems. First, it is hard to make sense of simple mental states, such as pains and tickles, other than as had by a subject or person. We can make little sense of an unowned or free-floating tickle. Second, a special problem is posed by more sophisticated mental states such as memory and intention. These states seem to have reference to persons built into their content. My memory of tasting coffee yesterday not only requires a current bearer but appears to implicate me in its content: I remember that I tasted coffee yesterday. Parfit is aware of
these problems and has an ingenious response, but it is fair to say that the view that mental states can be understood without reference to a person stands in need of considerable defense.

**Animalism**

What of the animalists’ answer? Let us be clear what that answer does and does not involve. Animalists do not hold that it is a necessary truth that all persons are human beings. Perhaps dolphins or chimpanzees qualify as persons. Perhaps there are nonhuman persons on other planets. What animalists do hold is that we (human persons) are human beings. That is, you are numerically identical to the human animal in your shoes.

A human being can survive in a coma, irretrievably devoid of mentality. According to animalism, if this fate befell me, I would continue to exist, but no person would then occupy my body. So I can exist without being a person, and hence, person (unlike human being) is not the concept of a fundamental kind of thing. Contrary to traditional views, the question of personal identity is not the same as the question of our identity, since we are not essentially persons.

**The Case for Animalism**

The animalist view strikes many as the merest common sense. Are we not flesh and blood creatures, members of the animal kingdom? Moreover, animalists have recently offered a compelling argument for their view. The form of the argument is this: Suppose animalism to be false and we end up with an absurdity, so animalism must be true.

Let us set up a version of the argument. Garrett is my name, the name of a person. We need a name of the human being in my shoes—call him Alf. The animalists’ claim is simple: Garrett = Alf. (Analogous identity claims hold for everyone else.) Suppose it is denied that Garrett = Alf. Then the following question arises: Garrett has mental states, but does Alf? Either answer to this question faces a significant obstacle. Suppose we say, “No, Garrett is a person and he has mental states, but Alf is just an animal, so he has no mental states only physical ones.” The trouble with this answer is that Alf has all my physical attributes, including all my brain states. It is now generally accepted that mental states, if not identical to brain states, are causally dependent on them. We have the mental states we have because of the brain states we are in. But in that case it seems unfair to deny mentality to Alf. He has a complex human brain (indeed the very brain I have), so why is he not the bearer of mental states?

Suppose we say, “Yes, fair enough, Alf does have mental states and is a person as much as Garrett.” But remember that the question we’re answering arose on the assumption that Garrett is not identical to Alf. So if Garrett is a person and it is now conceded that Alf is a person, then it follows that there are two people in my shoes! Generalizing from my case, it follows that the population of the planet is twice what we thought it was. But this is absurd. So to deny that Garrett = Alf leads to one of two absurdities: Either we are forced to fly in the face of the well-established thesis that mental states depend on, and are generated by, brain states or we are forced into a bizarre exercise in double counting. Hence, concludes the animalist, we should accept that Garrett = Alf.

**The Case Against Animalism**

This is a powerful argument. It is unclear exactly where it goes wrong, but go wrong it must, for its conclusion is manifestly false. If I were identical to Alf, then there could be no possible circumstance in which I survive but Alf does not or in which Alf survives but I do not. Yet there are such circumstances. Many would argue that, in the irreversible coma example, Alf survives but I do not.

There are more fanciful examples in which I survive yet Alf does not. As a highly trained athlete, I am naturally invited to be part of the first mission to Mars. The mission is a success and after a few months on Mars we return to Earth. After a routine checkup, doctors discover, despite no change to my appearance or mental life, that all my biological matter has been transformed into a hitherto unknown silicon-based life form. There is no animal (human or otherwise) in my shoes anymore! Alf is an animal, and essentially so: He could not survive without being an animal. Hence, Alf no longer exists. But I continue to exist. So I cannot be identical to Alf, contrary to animalism.

This reasoning might be thought sophistical. How can some bizarre and merely possible scenario in which I exist but Alf doesn’t have any bearing on whether I, here and now, am identical to Alf?
But here we must appreciate the logic of numerical identity. If \( A = B \) then everything true of \( A \) is true of \( B \) and vice versa. So if Garrett = Alf, then every possibility for Garrett must be a possibility for Alf and vice versa. Hence, if there is a possible scenario in which Garrett exists but Alf doesn’t (or vice versa) it follows that Garrett is not identical (here and now) with Alf.

**Concluding Remarks**

Our space travel thought experiment has done two things. It has undermined the animalists’ central claim: We are not identical with animals. It has also pointed us in the direction of a more promising theory of personal identity: the psychological view. Our thought experiment has made plausible the thesis that a continuing line of psychological continuity is sufficient for the continued existence of a person (whatever physical transformations he or she may have undergone). But there is a massive problem for this view. What if my stream of psychological continuity was to divide into two (e.g., if each of my brain hemispheres was transplanted into a new body)? The psychological view implies that I would be identical to the two resulting people, which is absurd.

Whether the psychological view can be defended against this worry, whether animalists can reply to the space travel objection, and whether Parfit can defend his Humean theory of the self are among the deepest contemporary questions in the metaphysics of persons.

_Brian Garrett_

**See also** Consciousness and Embodiment; Mind-Body Problem; Self-Consciousness

**Further Readings**


**PERSONAL IDENTITY, DEVELOPMENT OF**

Personal identity development has been defined in many ways over the past 50 years of social science history. Most social writers would agree, however, that one’s personal identity development is that which gives one a sense of purpose, meaning, continuity, and coherence in life. In the act of personal identity development, one finds expression for one’s own life meanings within a social context, and that context, in turn, provides recognition and mutual regulation of the individual and society. As Erik Erikson said in 1963, “For, indeed, in the social jungle of human existence there is no feeling of being alive without a sense of identity” (p. 130). This entry reviews key personal identity development concepts of Erikson and James Marcia. It also mentions four additional general approaches to understanding the development of personal identity.

**Views of Erik Erikson**

The concept of identity was first used and elaborated by psychoanalyst Erikson to describe a central disturbance among some young veterans returning from World War II. These men seemed to have lost a sense of inner sameness and continuity in their lives. Erikson thus began to refer to the concept of ego identity to describe a psychological entity that enables one to retain a sense of inner organization, sameness, and continuity across time and place—an entity under threat among his soldier patients. The psychoanalyst also stressed that one’s sense of ego identity development is dependent on the recognition and support that individuals receive from contexts meaningful to them—the immediate family, community, nation, and culture. The formation of an ego identity is thus dependent on the ways in which parents, teachers, social service providers in the immediate community, and representatives of the larger social structure meet and confirm individuals in their charge.

Ego identity is a product of the interaction between biological givens, psychological needs, and social forces according to Erikson. Thus, ego identity development is determined in part by one’s gender, physical attributes, strengths and limitations, in part by one’s conscious as well as unconscious needs, wishes, interests, and talents, and in part by
the roles and opportunities afforded one by the community. One's ego identity is, however, distinct from one's social roles. While the well-functioning individual may have many social roles in life, he or she has only one ego identity (unless there is serious psychopathology). The foundations of personal identity development begin in infancy through the images of and experiences with significant others whom one internalizes. Identity evolves during childhood based on the significant others with whom one emulates and tries to identify. However, in Erikson’s view, identity formation is more than the summation of all significant identifications of childhood; rather, identity formation is the sifting, sorting, and synthesizing of earlier important identifications into a new structure, greater than the sum of its parts.

Initial identity resolutions are generally undertaken during late adolescence, though identity formation and reformation remain lifelong processes, according to Erikson. It is during adolescence that the biological changes of puberty, alongside one's growing capacities for pursuing psychological interests and values, in combination with societal demands for the assumption of adult roles and values that personal identity concerns, often first come to a head. He described the main psychosocial undertaking of adolescence to be that of finding an optimal resolution to the identity versus role confusion task. A time of identity exploration and experimentation are vital to optimal identity formation. Failure to undergo this identity formation process will leave the individual either drifting and centerless, an uninvolved spectator in life, or oppositional and antagonistic, devising an identity based on those values that parents would hate most, according to Erikson. Optimal identity formation serves as the cornerstone to the eventual expression of intimacy, both with friends as well as a life partner.

The Identity Statuses
Since Erikson’s original writings on ego identity, many writers have either extended his work or reformulated the meaning of identity from a somewhat different perspective. One of the writers most noted for elaborating and extending Erikson’s concept of identity has been James Marcia. Marcia suggested that ego identity is better conceptualized according to one of four basic styles (or identity statuses) by which individuals seek (or not) resolutions to questions of personal identity. There are two styles of commitment: identity achievement and foreclosure. The identity achieved individual, as per Erikson's definition, has undergone a period of exploration and experimentation prior to making meaningful identity-defining commitments on his or her own terms. The foreclosed individual has also adopted firm, identity-defining commitments but has done so without prior exploration; the identity-defining commitments of the foreclosed are based on identifications with significant others, particularly parents. Similarly, there are two styles of identity noncommitment: moratorium and diffusion. The moratorium individual is in the process of searching for meaningful identity commitments, while the diffuse individual is uninterested in making such commitments. These four identity statuses have been differentially linked with various attachment styles, patterns of family communication, personality variables, behavioral consequences, and developmental patterns of change.

Additional Theoretical Approaches
Other general approaches to defining personal identity include the following: (a) the narrative approach, in which writers suggest that personal identity does not exist until one constructs and tells a story about the self; (b) the sociocultural approach, wherein writers emphasize the roles of culture, society, and the media as primary forces in defining individual identity; (c) the structural-stage approach, in which intrapsychic elements are emphasized in the individual identity formation process; and (d) an historical perspective, in which changing historical conditions are regarded as primary in regulating personal identity development.

Jane Kroger

See also Moral Development; Relationships, Development of; Self, Development of

Further Readings
The division of responsibility for a given behavior between the individual and the situation is central to much work and debate in psychology. This distinction is the foundation of attribution theory, underlies the division of personality from social psychology, and bears on debates about the consistency of personality across situations. The division in many ways resembles the debate between nature and nurture, though both of these, being aspects of the individual, fall into the personality camp. Also as in that debate, there has been a growing sense of the inadequacy and artificiality of the opposition of person and situation and increased emphasis on an interactionist perspective. This view of the interplay of aspects of the person and the situation in producing behavior is to some extent an empirical issue, based on data on the limited predictive power of pure person and situation accounts, and also partly a theoretical issue, based on a closer examination of what is meant by explanation and causality.

**Internal Versus External Attributions**

Attribution theory, in its most basic form, provides the rules by which people distinguish, or ought to distinguish, between behaviors that are caused by attributes of the individual versus those that are the result of the situation. When a woman is observed gesturing rudely to another motorist, one can inquire whether that indicates a short-tempered hostility in her or whether instead being cut off on the highway represents an intolerable provocation. In Harold Kelley’s classic formulation, one would gather information on whether she consistently gestures in that manner, whether she is alone in behaving that way, and whether, in other situations, she displays such behavior. To the extent that such questions are answered affirmatively, the casual responsibility is attributed to her personality and not to the situation.

The attribution framework not only specifies the process of assigning causality, but it also, more practically, provides a way of formulating predictions. If the attribution is internal, attributed to something about the gesturing driver, then one can predict that he or she is more likely to show that behavior in other settings, while if it is external, attributed to the situation, then one can predict that other people are also likely to exhibit the same behavior in a similar situation. In this way, the distinction between person and situation is connected to two other notions: the stability of individual differences and the power of the situation.

**Personality Consistency**

Questions of the stability of personality, however, have troubled the study of individual differences for some time. While people are stable across time in the same setting, they appear quite inconsistent across situations. People who are aggressive in one setting will, in another setting, be quite meek, and those who are honest in one domain may be quite duplicitous in another. In fact, such findings led Walter Mischel to suggest, in a controversial book that still frames much of the debate in the field, that personality consistency correlations across situations rarely exceed .3 and that the search for broad predictive traits is unlikely to prove fruitful. Those consistencies that are found are frequently a result of method variance or self-report bias. People who say they are generally honest also report that they do not often lie; it is when specific concrete behaviors are measured that consistency proves elusive.

Efforts to overcome the difficulty of predicting specific behaviors have taken several forms. One approach has been to suggest that it is a matter of reliability and that aggregating over multiple instances of concrete behavior will make prediction possible. Another approach has been to move from nomothetic approaches, which should apply to all people, to more ideographic ones, suggesting that for only a subset of the population—for example, those to whom honesty is an important dimension—will there be consistency from one setting to another. Another approach is to suggest that superficial traits may be unstable, but that central ones, such as the “big five,” form a stable core to personality. Such...
Perspective Taking in Language Processing

approaches, while they have made some headway, have not come close to establishing a way of characterizing broad consistencies in behavior.

**Interactionism**

What has been an effective approach is to incorporate the person-by-situation interaction into the model. The critical aspect of this approach is not to expect consistency in behavior across all situations but instead only across ones that share essential features. One might find, for example, that a person is consistently shy in situations involving groups of older people but that this trait is not manifest with peers or one-on-one. In essence, it is a way to find clusters of situations that are psychologically similar and so tap the same aspects of personality. Or viewed another way, it allows one to predict behavior based on consistency over time in what are effectively similar situations.

This limited view of the power of personality in behavior appears to be addressing the same question as that addressed by attribution theory. It also seems consistent with conclusions about the power of the situation from such findings as those of Stanley Milgram's classic experiment. In this work, perfectly normal people are led, by the situation and the instructions of the impassive experimenter, to administer apparently lethal shocks to an innocent stranger, regardless of their own personality. However, the notion of limited generalizability of personality across situations is, in important ways, orthogonal to those issues. The exquisite sensitivity of personality to nuances of the situation makes it hard to predict behavior from personality, but it does not mean that personality plays any smaller a role. Nor does the narrowness of the expression of traits mean that situations are powerful. The variability in behavior from one person to another, in a given situation, is a result of individual differences, regardless of whether they are predictable or stable. That is, if one person cries and another does not at the end of a tragic movie, there is no question of person versus situation as an explanation, and the debate is just about in what other situations he will cry.

This issue is also captured with a closer look at attribution theory. If one asks whether that crying behavior is due to the movie or the situation, one needs first to specify the implied comparison group. His crying more than other people is due to something about him; his crying more during the movie than while waiting in line for popcorn is a result of the situation. To suggest that either the person or the situation is generally more causally powerful is not a meaningful question. One can view the strength of personality as being the degree to which individual differences in a situation are predictive. The strength of the situation can be viewed as the degree to which the average behavior deviates from what one might intuitively expect in such a situation (as with the Milgram experiment) or from behavior in an apparently similar setting (as is shown in findings of limited generalizability of dispositions). Thus, the strength of either person or situation has no bearing on the weakness of the other. Nonetheless, it does seem to be the case that it is harder to predict from personality than most people, including psychologists, expect and also that small changes in the situation have a more profound effect than most expect, and so there is some meaning, albeit an ambiguous one, to the suggestion that the person is weak and situations strong.

Nicholas J. S. Christenfeld

See also Attitudes and Behavior; Attribution Theory; Character and Personality, Philosophical Perspectives

**Further Readings**


**Perspective Taking in Language Processing**

Perspective taking is the spontaneous consideration of another's mental states—thoughts, beliefs, and goals—to understand how they interpret a given situation. The ability to appreciate similarities and differences in perspective is important whenever it is necessary to coordinate one’s own actions with those of other people. Such coordination problems arise in many different social activities, such as in economic games involving cooperation and competition. Perspective taking
has thus been studied extensively by cognitive and social psychologists, linguists, anthropologists, computer scientists, and economists.

Participating in a conversation requires coordination of thought and action on many different levels. Language users must coordinate their thoughts and actions because language is ambiguous: There is no one-to-one mapping between what people say and what they mean. A critical question in modern psycholinguistics is the extent to which perspective taking is necessary for coordination to succeed. The audience design hypothesis assumes that people speak and understand vis-à-vis their common ground: the set of information that they believe to be mutually shared. Under this view, speakers should take their audience’s informational needs into account when deciding what to say. Similarly, listeners should consider speakers’ knowledge when interpreting what they mean. Many theories of language use assume that perspective taking plays a fundamental role in speaking and understanding. However, psycholinguistic studies suggest that perspective taking may play a more limited role than previously believed.

**Perspective Taking by Speakers**

Experiments show that speakers include information in their speech that is obviously crucial for understanding. For example, when they identify a building to someone from out of town, they are more likely to use a description than a name. But in numerous cases where speakers could phrase their speech in ways that would reduce ambiguity for the addressee, they often do not. They may call a baseball bat “the bat,” even when the listener could think that they are talking about the animal, a bat. They choose among different sentence structures based on what is easy or hard to produce, not on what would be easy or hard for the listener to understand. Although certain sentence structures can be made unambiguous for the listener by using certain forms of phrasing, emphasis, or intonation, the evidence again suggests that speakers do not do so. Similarly, speakers tend to pronounce a word more clearly when it is less predictable, but this depends more on how recently the speaker has produced the word, not on the listener’s need to understand.

When speakers do tailor their speech to a listener’s needs, it seems to require much conscious attention and active monitoring for potential ambiguities. Consequently, when speakers are cognitively busy, such as when they are multitasking, their ability to tailor their speech to the listener is drastically reduced.

To design what they say for their addressees, speakers also need to be able to interpret their utterances from the perspective of the addressee. But speakers encounter major difficulty doing so. For example, they do not fully take into account the constraints of the medium. Utterances are more ambiguous on email than when spoken, but people expect addressees to understand them to the same degree. Furthermore, when people try to use intonation to disambiguate what they say, knowledge of their own intention makes the utterance seem less ambiguous than it really is. Consequently, when they try to take the perspective of the addressee, they routinely overestimate how well the addressee will understand them. Speakers even have the same difficulty taking the perspective of their future selves, overestimating their own ability to understand what they meant when they listen to a recording of themselves in the future. This idea was foreshadowed in the 16th century by Michel de Montaigne, who wrote that when reading his own writing he could not always find his original thought: “I do not know what I meant to say, and often I get burned by correcting and putting in a new meaning, because I have lost the first one, which was better” (pp. 425–426).

**Perspective Taking by Listeners and Readers**

Although listeners do consider the perspective of the speaker when they interpret speech, their ability to do so is limited. For instance, when asked for help, listeners interpret the request based on the physical constraints of the speaker. Listeners do expect speakers to design what they say for them, and they attempt to take the speaker’s perspective. But it is currently a matter of debate just how completely listeners are able to use this information to constrain how they process what is said. Under some circumstances, listeners appear able to use this information during the early moments of the processing of a word or phrase. Even young children seem to use perspective information in this way. However, experiments that have more closely examined the time course of language processing suggest a different story: Listeners may be able to take the speaker’s perspective into account to anticipate what the speaker might refer to, but they cannot immediately integrate this information when they process what the speaker actually says.
Occasionally, listeners show complete disregard of the speaker’s perspective. This results in a surprising degree of egocentric-based errors when the perspectives of the listener and the speaker diverge. Such perspective errors are more prevalent in young children but are also present with adults. This suggests a developmental continuum in which listeners learn over time to make better use of the speaker’s perspective but not to completely incorporate it into the comprehension process.

Readers of literary works may also adopt the perspective of protagonists when they interpret what the protagonists say or how they would understand language. Yet when studies unconfound the perspective of the reader from the perspective of the protagonist, they discover that readers interpret text from their own perspective. It is later in the process of understanding that attempts are made to allow for the perspective of the protagonist, if it is different from their own.

**Perspective Taking and Successful Communication**

The idea that perspective taking plays a limited role in the coordination of meaning may be seen as conflicting with the intuition that speakers and listeners are routinely successful in communicating their intentions. Instead, it may indicate that the coordination of meaning proceeds via other mechanisms. For instance, there is evidence that a community of agents can coordinate meaning without a representation of any agent’s perspective. To conclude, people sometimes use perspective to coordinate meaning, but they need not do so to successfully communicate. During a conversation people’s perspectives typically come into alignment, thereby making the active consideration of the other’s perspective superfluous.

*Boaz Keysar and Dale J. Barr*

---

**Persuasion**

The way people change their attitudes, beliefs, and behavior is often influenced by their experiences and their interactions with others. The mechanism through which people adopt and change their attitudes and behaviors based on contextual factors is called social influence. Robert Cialdini defines social influence as a change in one’s attitudes, beliefs, or behaviors, which may be due to real or imagined external pressure(s). Typically people are exposed to a number of influence attempts daily, sometimes without being consciously aware of it. On any given day, a person may be prompted to apply for a credit card, buy a product advertised on television or radio, and change his or her opinion about a politician by reading a news article online. Persuasion is one component of social influence. Specifically, persuasion is the process by which an individual is influenced to adopt or change a particular attitude or belief. People can be persuaded through a number of different avenues (other people, advertisements, books, TV, Internet, etc.).

This entry focuses on the influence of persuasive messages on peoples’ attitudes and beliefs by reviewing the prevailing theoretical models of persuasion. These dual process models indicate that when forming or changing an attitude, people either think deeply about the content of a persuasive message or make decisions based on surface characteristics associated with the message, such as the attractiveness of the individual trying to persuade them. This entry also reviews general factors that may affect message processing, and the six principles of influence.

**Key Definitions**

As stated above, social influence refers to the change in one’s attitudes, behavior, or beliefs because of
external pressure that is real or imagined. There are two general types of social influence: persuasion and compliance. Compliance research examines changes in behavior resulting from a direct request. For instance, if an individual is asked to sign a petition advocating that the U.S. government adopt a universal health care system and he or she agrees to this request, the individual is complying with this request. Persuasion is focused on the change in a private attitude or belief as a result of receiving a message. So, for instance, if an individual reads a newspaper editorial that contains compelling reasons why the government should adopt a universal health care plan and the individual’s opinion on this topic is changed as a result of reading the arguments, the individual has been persuaded.

Additional terms relevant to an understanding of persuasion describe the individuals involved in an influence attempt. First, influence practitioner, communicator, or agent of influence are terms used to describe the individual who attempts to influence others. For instance, in the example above, the person who made the request to sign the petition to change the health care system and the person who wrote the editorial are the influence practitioners. Next, the target or target of influence refers to the person who has the influence attempt directed at him or her.

Dual Process Models of Persuasion

The prevailing theoretical models of persuasion are the elaboration likelihood model (ELM) and the heuristic systematic model (HSM). These complimentary models explain how individuals process persuasive information and make predictions about how that information influences an individual’s own attitudes or beliefs.

The Elaboration Likelihood Model

Richard Petty and John Cacioppo introduced the ELM in 1981. The ELM proposes that there are two routes through which attitude change occurs: a central route or a peripheral route. Central processing occurs when an individual thinks carefully about a persuasive message. That is, a person will carefully consider the persuasive message to determine how much merit the argument(s) have. In this case, higher elaboration is involved with processing the message and any resulting change in attitudes or beliefs tend to be long lasting. In other situations, people are more likely to process a persuasive message through the peripheral route. Peripheral processing occurs when an individual pays more attention to nonmessage relevant factors, such as attractiveness of the influence agent or the quantity of his or her persuasive arguments. In this case, any resulting changes in attitudes or beliefs tend to be short term because the influence target is not considering the central theme of the message.

In their 1984 experiment, Petty and Cacioppo found evidence for these two different routes to persuasion. They reported that when an issue under consideration would directly impact the target of persuasion, individuals were more likely to think about and be persuaded by stronger arguments and by larger numbers of arguments, indicating central processing of the message. However, when the issue was not directly relevant to the target of persuasion, arguments were not considered as carefully and the number of arguments rather than the quality was more persuasive. This is because the message was peripherally processed and the number of arguments, rather than argument quality, served as a cue of message validity.

The Heuristic Systematic Model

In 1980, Shelly Chaiken proposed a dual processing model of persuasion known as the HSM. Similar to the ELM, the model proposes two routes to persuasion: systematic processing or heuristic processing. However, unlike the ELM, the HSM proposes that it is possible for message processing to occur simultaneously through both routes. According to the HSM, it can be both the quality of the arguments and the quantity that simultaneously provide information relevant to persuasion. When systematically processing a persuasive message, a person engages in more effortful thinking, scrutinizing the message carefully and engaging in more cognitive effort. Heuristic processing is less thoughtful processing in which individuals rely more on cues, such as the likeability of the influence agent to determine the validity of the message.

In one of her 1980 experiments, Chaiken demonstrated the different message processing routes to persuasion. Similar to Petty and Cacioppo, she found that when an individual had more of a personal stake with an issue, they were more likely to systematically process a persuasive message. On
the other hand, when issue involvement was low because they had no stake in the issue, influence targets were more likely to use heuristic processing. Under conditions of high issue involvement, people showed more opinion change when there were more arguments, suggesting more thoughtful and systematic processing. However, when an issue was less relevant, persuasion was impacted more by likeability of the influence agent rather than the number of arguments, indicating a reliance on more heuristic processing and use of cues. Furthermore, individuals in the study who were more involved with the issue spent more time thinking about the arguments, recalled more arguments, and generated more message-relevant thoughts.

**Additional Factors Impacting Persuasion**

Over the years, research on persuasion has examined many different factors that affect which route to persuasion a person will use to process a persuasive message. Some factors that have been consistently found to affect the processing of persuasive messages and the overall effectiveness of the persuasion attempt include issue importance or personal relevance (as noted above), source credibility and attractiveness, motivation and ability, and the strength or quality of the argument(s) in the message.

Specifically, individuals process persuasive messages differently depending on whether the source of the message—the influence agent—is someone they perceive to be credible. For instance, an individual may be persuaded to adopt a more healthy diet based on the advice of his or her doctor but may not have been persuaded based solely on the advice of a friend. This is because doctors are assumed to have more expertise in health-related knowledge. In this case, people rely more on heuristic processing, by using the expertise and status of the communicator to evaluate the message. Similarly, attractiveness is another cue people often use as a peripheral or heuristic cue when evaluating a persuasive message. That is, people may judge a persuasive message to be more valid if the source of the message is attractive because people generally associate positive characteristics with attractive people.

As noted above, when a message has more personal relevance or when people are more highly involved with it (i.e., when it may directly impact them), people are more inclined to process the message centrally or systematically. For example, someone who cares deeply about animal rights may be more inclined to think effortfully about a potential law involving the treatment of animals. Similarly, a person will more carefully consider voting for a law banning after hours operation of an airport if their home is near the airport.

Motivation and ability also play a significant role in determining how a message will be processed and subsequent persuasion. Under certain conditions a person may vary in the cognitive ability or the motivation to carefully process a message. For instance, a person who has spent a week studying hard for a biology exam may not be able or motivated to thoughtfully process a message advocating a new university-wide security policy. Thus, he or she may resort to peripheral or heuristic processing. A person who has had a relatively stress-free week and no exams may be more likely to elaborate carefully on the new security policy, especially if he or she believes the proposed policy will impact his or her life, because both motivation and ability will be high. Finally, the quality and quantity of the persuasive arguments may also influence persuasion. In general, research indicates that stronger, higher quality arguments are more persuasive than weaker, poorer quality arguments.

Overall, there is consensus in the literature on persuasion that a combination of these factors work together to influence the processing of persuasive messages. Generally, when a person has high motivation and ability to process, the argument is strong, and the issue is relevant, more elaboration will occur and this will produce an enduring shift in attitude. On the other hand, if motivation and ability to process are low, people may rely more on cues such as source attractiveness or likeability to make a judgment that will likely produce short-term persuasion.

**The Six Principles of Influence**

While people are capable of processing persuasive messages using the different routes reviewed above, people tend to be cognitive misers. That is, people process information using heuristics or rules of thumb that help simplify decision making. Robert Cialdini proposes that most influence targets respond to a set of trigger features for persuasion and compliance. Specifically, he argues that there are six common principles—heuristics or short cut
rules—that people are likely to use when confronted with a persuasion attempt by an influence agent. Responding heuristically allows individuals to react quickly, saving time and mental energy. Heuristic responding is not only efficient, but it allows individuals to make informative and accurate judgments when mentally overloaded. As a result such automatic processing can often lead individuals to make accurate decisions. These six principles of persuasion are as follows: authority, liking or friendship, scarcity, social validation, reciprocity, and commitment or consistency. This entry will review each of these principles below.

Authority

When relying on heuristics, people tend to be more persuaded by an influence agent if they perceive the source as an authority figure. The authority heuristic indicates that if an expert says it’s true, then it must be. One study on the influence of an authority indicated that people were 3.5 times more likely to follow a jaywalker in a suit than one in casual clothing. Other research indicates that people heed the recommendations of experts such as doctors, scientists, or executives because these are authority figures that are assumed to have more knowledge than laypersons. This aspect of the authority heuristic also often appears in advertisements that feature product endorsements from actors who portray doctors on television shows. Thus, targets of influence will be influenced by agents who appear to be authority figures.

Liking and Friendship

Research on the influence of liking on persuasion indicates that under most circumstances, people are more persuaded by individuals that they like or find attractive. This is typically based on a second heuristic involving the reliance on likeable individuals as good sources of information. This liking heuristic indicates that if a likeable person (especially if he or she is also similar) endorses something, it must be good. For instance, dating or wardrobe advice from a friend will likely be more influential than similar advice from an acquaintance or stranger. Specific research on this question indicates that negotiators who find common goals and shared interests with the opposing negotiator will be more successful in finding mutually beneficial outcomes. Additionally, celebrities make effective agents of persuasion because they tend to be both well liked and also perceived as experts.

Scarcity

Another heuristic cue involves scarcity. The scarcity heuristic states that if it is rare, it must be good. That is, if something is not widely available, then it is perceived as valuable. For example, during the 1990s Beanie Babies became a popular stuffed animal. Many of these stuffed animals were rare and limited in production. This scarcity led to an increase in prices, ultimately leading to a higher demand for the toys. People came to believe that this product was more valuable and increased their desire to have the product. Because of the scarcity of the plush toy, an item that cost around $0.25 to produce sometimes sold for hundreds of dollars.

One study illustrated the importance of scarcity in the context of sales of Australian beef. A company selling beef knew there would be a shortage of a certain type of beef from Australia. To examine the impact of this scarcity, the researchers created 3 different sales scripts. With the standard script, customer service representatives called customers and took their orders as usual with no mention of the upcoming shortage. With the scarcity script, the representatives called customers and took their orders while mentioning the upcoming shortage. With the scarcity and exclusivity of information script, the representatives called customers and took their orders while mentioning the upcoming shortage. With the scarcity script, the representatives called customers and took their orders while mentioning the upcoming shortage. With the scarcity and exclusivity of information script, the representatives called customers and took their orders while mentioning the upcoming shortage. The average amount of beef ordered using each version of the script illustrates the persuasive impact of scarcity. With the standard script, an average of 10 loads of beef was ordered; with the scarcity script, an average of 24 loads of beef was ordered; and with the scarcity and exclusivity of information script, an average of 61 loads of beef was ordered. Thus, these results indicate that scarcity in terms of both product supply and information is very persuasive.

Social Validation

Social validation or social proof is the tendency for people to look to others to determine appropriate attitudes or behaviors. Across cultures and situations, people follow social norms or rules for
behavior and change their behavior to match the actions of others. This phenomenon is enhanced in ambiguous situations where an individual is unsure of the appropriate response. For example, one study examined the extent to which people are influenced by the actions of others by varying the number of people looking up at nothing and observing the number of passersby who also stopped to look up. The researchers placed either one person or five people on a busy street looking up and staring at nothing. They found that 5% of passersby stopped and looked up too when only one person was looking up but 80% of passersby looked up when it was the group of five.

Reciprocity

The rule of reciprocity states that people are obliged to give back to others what they have given to us. People are more likely to be persuaded if they feel they owe the influence agent a favor. For example, if a person asks some of his or her coworkers to help him or her move, the coworkers that had called on this person to help them move in the past will likely be the ones to volunteer. Reciprocity is influential both within individuals’ social networks—research indicates that widowed women who gave and received equal amounts of emotional support were happier than those who either solely gave or received too much—and outside them too—a study that examined charitable donation rates reported that the inclusion of “free” address labels from the charity increased donation rates from 18% to 35%.

Commitment and Consistency

People are creatures of habit. The last principle, commitment and consistency, capitalizes on tendency. Generally, people will look to previous attitudes and behaviors when confronted with an influence attempt. Individuals will act or think in accordance with previous actions or thoughts. For example, if a person was previously gregarious at a party, he or she may act the same way at another party to remain consistent with past behavior. Furthermore, if a person makes a commitment to a certain behavior or idea, he or she will be more likely to be persuaded to commitment to a related request at a later time. A well-studied example of this is called the foot-in-the-door effect, which was introduced in 1966 by Jonathan Freedman and Scott Fraser. In one experiment, housewives who were contacted via telephone and asked to complete a short survey on household products were more likely to later agree to allow men to come into their house and classify all their household products. Those housewives who were not first asked to participate in the telephone survey were far less likely to comply with the request to classify household products.

Overall, these six principles of influence are widely used cognitive shortcuts that people rely on to efficiently respond to social influence attempts. Since persuasive attempts may be directed at individuals in a quick and unexpected manner, this use of peripheral or heuristic processing can allow an individual to process and respond to the persuasive communication in a timely manner. Such heuristics or cues may lead to changes in one’s attitudes and also in one’s behaviors. However, relying on cues such as these principles can lead to errors in message processing as well. For instance, by relying on heuristics, one may come to reject a strong and valid argument in favor of one that comes from a well-liked source.

Rosanna E. Guadagno and Nicole L. Muscanell

See also Attitude Change; Attitudes and Behavior; Political Psychology; Social Cognition

Further Readings


**PHENOMENOLOGY OF ACTION**

This entry concerns the consciousness or experience subjects have of their own actions. The now burgeoning interest in this phenomenology of agency is a fairly recent development, even more recent than the rediscovery of consciousness in general. This is reflected in the fact that much of the literature, especially the philosophical literature, is concerned with defending or battling viewpoints still skeptical of the significance of a distinctive phenomenology of agency or even of its very existence. One source for such skepticism is the assumption that such a phenomenology would have to take the form of a specific and unitary feeling or sensation of acting. A better conception is that of a family of actional experiences. Consider the following everyday scenario: You plan to write a paper, and after much deliberation, you choose a topic and create a rough outline. Sitting down in front of your computer to write, perceiving its screen as something to be filled and the keyboard as a means to this end, you focus your thoughts on creating a sentence, finally executing typing movements, experiencing yourself as moving your hands and moving the keys through them and perceiving the events of letters appearing on the screen as the result of your movements. These are some of the family of experiences connected to action and its authorship: experiences of deliberating, intending, of active and purposive bodily movement, of perceiving entities as objects for and results of action. They raise a host of questions. This entry focuses on the structure of actional experience in the sense of the experience of active, purposive movement, and its relation to perceptual experience.

**The Structure of Actional Experience**

Attempts to find some order in action consciousness in recent times have often taken the form of a classification of different kinds of intentions or other action representations. John Searle influentially distinguished “intentions in action” from “prior intentions”—the intentions accompanying action (“I am doing A”) from those preceding it (“I will do A”). Other authors have proposed related distinctions, for example, between proximal and distal intentions. But contrary to what these terminologies suggest, the difference between, say, the prior intention of raising one’s arm and the experience of actually raising it is not solely temporal. It is also structural: They differ in their representational format. These differences can be fruitfully described using the notion of nonconceptual content found in the theory of perception. The nonconceptual content of actional experience is nonsymbolic and richer, more fine-grained and denser than the symbolic, conceptual content of prior intentions.

The distinctions between these action representations are also closely related to different levels of control and ownership of actions and to the corresponding failures, slips, and pathologies. For example, a person may control and own a habitual action at the nonconceptual bodily level but not at the conceptual, rational level of deliberation. A related pathological phenomenon is utilization behavior, in which patients are unable to inhibit certain stereotypical actions such as drinking from a cup placed in front of them.

**Actional and Perceptual Experience**

Some authors claim that actional experience and perceptual experience are more than just structurally similar. They suggest that actional experience is simply a species of perceptual consciousness, even that there is a sense of agency, as there are senses of touch and sight. Searle defends a contrary view, according to which actional consciousness is diametrically opposed to perceptual consciousness in one fundamental respect. Perceptual experience is essentially passive: In perception, we experience ourselves as achieving fit between mind and world by being receptive to an independently existing world (mind-to-world direction of fit). By contrast, actional experience is active: In action, we experience ourselves as making something happen, as achieving fit by adapting the world to the contents of our action representations (world-to-mind direction of fit). It is a condition of adequacy on the perceptual theory that it must account for this fundamental phenomenological difference between action and perception. If this difference is reconstructed as a difference between two basic kinds of perceptual experience, in which we experience ourselves in an active or, respectively, in a passive causal role, it may end up being merely verbally different from Searle’s view.

Another difference between the perceptual account and Searle’s account is that Searle treats the representational content of actional experience as constitutive of action itself. Action occurs when the content of actional experience is satisfied, when the bodily movement I experience myself as bringing
about actually occurs. In this respect, actional and perceptual experience are treated in exactly parallel fashion despite their fundamental difference, as perception is also constituted by satisfied, veridical, perceptual experience. In contrast, the perceptual account treats action as independent of the experience of acting, which is denied a constitutive role.

Whether actional experience is perceptual or not, there is a wealth of empirical data on the relation between these kinds of experience. It is tempting to think that proprioceptive experience is a prerequisite for the experience of action. But the famous case of Ian Waterman, a patient who lost all sensory input below the neck but still experiences himself as an agent, seems to show that this is not so. Indeed, it has been suggested that the attenuation of proprioceptive experience rather indicates the presence of action. Many studies show that the proprioceptive consequences of one’s own actions are attenuated—this is why one can’t tickle oneself. By comparison, exteroceptively accessed effects of one’s purposive movements are bound together with these movements—for example, a sound and the bodily movement that produces it are experienced to be temporally closer than they actually are.

Michael Schmitz

See also Action and Bodily Movement; Philosophy of Action

Further Readings


PHILOSOPHY OF ACTION

The use of the label *philosophy of action* gained currency roughly in the second half of the 20th century. However, although the label is relatively new, the subject matter is not. At least since Socrates, philosophers have been concerned with the problems and questions now gathered under that label. Essentially, the philosophy of action seeks to offer an account of distinctively human behavior—in particular, of behavior that is characteristic of, to use Aristotle’s phrase, “rational animals.” This is behavior on the basis of which we make judgments about people’s goals, characters, and values, and it is the behavior that grounds ascriptions of causal and sometimes also moral and legal responsibility to people for certain occurrences, outcomes, and states of affairs.

A philosophical account of the behavior just mentioned requires a good understanding of issues such as what exactly actions, and their counterparts—omissions—are; when an action is voluntary and intentional; whether there are genuinely free actions or whether freedom is an illusion; and, if there are free actions, what roles do reasons, intentions, and the will play in such actions (which introduces the problem of *akrasia*—the possibility of acting against one’s better judgment because of “weakness of the will”). Other issues include how we should understand the explanation of actions by reference to reasons, the role of emotions and of the unconscious in actions, whether moral responsibility for actions (and omissions and their consequences) requires free agency, how to understand collective agency, and the agency of non-human animals.

One distinctive feature of the philosophy of action is that its boundaries are relatively vague. The reason for this is that the core questions in this area of philosophy cannot be addressed without resolving problems in other areas, such as metaphysics, the philosophy of mind, the philosophy of language, ethics, and legal philosophy. Throughout its history, the agenda of the philosophy of action has been defined by different questions, and, as a result, the central debates have, at different times, been closely tied to different areas in philosophy. This entry provides an overview of some of the central questions and arguments in the philosophy of action, as well as an indication of which issues have been at the center of contemporary philosophy of action.

From Aristotle to the 20th Century

Aristotle is probably the philosopher who dealt most thoroughly with the various issues in the philosophy of action. Throughout the centuries, most of the
great philosophers grappled with the problems he raised and introduced new ones, but it seems right to say that the 20th century saw a revival of interest in this area of philosophy. This was partly due to the publication in 1957 of *Intention*, a deceptively short book by Elizabeth Anscombe, a disciple of Ludwig Wittgenstein’s, who was also much influenced by Aristotle and by St. Thomas Aquinas. *Intention* was greatly influential, and although many of the views she defended there have been forgotten or were never embraced, many of the contemporary debates on action are framed in relation to her treatment of the subject. Among other things, her discussion placed questions about intentional actions, which she characterized, roughly, as actions done for a reason, center stage—displacing questions about voluntary actions, which the tradition she was writing against had characterized as actions caused by *volitions* (acts of will) and which had been the staple of earlier discussions. In the years following the publication of *Intention* and especially after the publication in 1963 of Donald Davidson’s paper “Actions, Reasons, and Causes,” the theory of action concentrated mainly on questions about the metaphysics of actions and about the relation between actions and reasons in the production and explanation of intentional action. More recent work has focused on debates about free will and moral responsibility, autonomy and control, reasons and rationality, and knowledge and action, among other things.

The Metaphysics and Explanation of Actions

As mentioned above, a central question in this field is what actions are, together with related problems concerning the individuation and spatiotemporal location of actions. In *Intention*, Anscombe emphasizes the fact that agents, whether intentionally or not, often do one thing in, or by, doing another. For instance, I may replenish the water supply of a house by pumping, which I may do by moving my arm up and down. So, in cases when one does one thing by doing another, we need to decide whether we are dealing with one particular action (perhaps amenable to various descriptions) or with many. And if many, we may wonder how these actions are related: As part and whole? As cause and effect? In some other way?

Since the mid-1980s, the dominant doctrine has been that an action is an event (a bodily movement), caused and/or causally explained by a reason (a combination of mental states) that rationalizes the action when the latter is considered under the right description—the description “under which” the action is intentional. An action is also the cause of other events (the action’s effects), on account of which new descriptions can be applied to the original (basic) action. Accordingly, when someone does one thing by doing another, there is only one event—an action—amenable to many descriptions. Thus, on this picture, my action of pumping water is an event (the motion of my arm) caused by my desire to replenish the water supply and my belief that pumping water is the way to do so. The basic action, the motion of my arm, is an event that can be redescribed as my pumping water, as my replenishing the water supply, and so on, because that action event causes further events: the motion of the water along the pipes, the filling up of the water tank, and so on.

This doctrine about the relation between actions and reasons, known as the *causal theory of action*, was articulated by Donald Davidson. It is still probably accepted by most philosophers even if some disagree on details and even though its detractors have emphasized its various shortcomings.

One of these shortcomings is that accounts based on this picture tend to deal with omissions mostly as an afterthought. But omissions are an important part of the behavior of agents that falls within the province of the philosophy of action, not least because omissions and their outcomes are susceptible to questions about causal and moral responsibility and also because agents often omit to do things intentionally and for a reason. A central question relating to the distinction between actions and omissions is whether there is, or there must always be, any morally significant difference between them, or between “making something happen” and “allowing it to happen”—for instance, between killing and letting die. Another is that of the relation between causal and moral responsibility. It is often assumed that direct causal responsibility for an event is a necessary condition for someone to be held morally responsible for that event. But omissions seem to undermine that thought. For it seems that an agent might be morally responsible for an event that she did not contribute to causing if by causing we mean that some motion of her body caused, however indirectly, that event—for example, when the agent
failed to prevent the event in question through an omission. In other words, it seems that, when certain conditions obtain, just allowing an event to happen could justify an attribution of moral responsibility for that event to the agent that allowed it to happen. (An example might be someone who allows a child to drown, knowing that he could have saved the child without danger to himself, even when he was not causally or morally responsible for the child’s predicament.)

Other difficulties associated with Davidson’s view concern the idea that actions are events. For instance, the idea that a basic action is to be identified with an event that is a motion of one’s body seems to generate problems: To name one, if the causing of the motion (my moving of my arm) and the motion caused (the motion of my arm) are identical, it seems to follow that the causing of an event is identical to the event thus caused, which seems absurd. Moreover, although it seems easy to give some actions a precise location in space and time (e.g., my action of opening a tap), this is not so for other actions: Where and when are we to locate John’s action of killing James, when he kills him by shooting him on Monday in the park, but James dies on Tuesday in a hospital? The same does not seem true of ordinary (nonaction) events.

The causal theory also faces the so-called problems of deviant causal chains and of the irrelevancy of the mental in the causation of action. Davidson himself diagnosed the problem of deviant causal chains, for he realized that invoking causation, if necessary, is certainly not sufficient for a satisfactory account of what it is for an agent to act for a reason. The problem is that an action might be caused by a reason that rationalizes it without its being true that the agent acted for that reason. Davidson’s own example is of a climber in danger who wants to rid himself of the weight of his companion and believes that he can do so by loosening his grip on the rope, and where these very considerations so unnerv him that they cause him to loosen his grip. So his belief and desire cause him to loosen his grip and yet he does not loosen it for the reason he had for doing so: He doesn’t loosen his grip for any reason. This suggests that a reason must not just cause the action but must cause it in the right way. However, there is no widely accepted account of what this right way is.

The difficulty generated by the threat of the irrelevancy of mental can be expressed as a dilemma: In the causal theory, either actions are causally overdetermined, or the mental is irrelevant or epiphenomenal. According to the standard causal theory, actions are events caused by reasons, which are mental events (or states). But actions are typically conceived of as motions of the body and hence as physiological events, and if so, these motions are, we are told by neurophysiologists, caused by other physiological and neural events. So either reasons are identical to the neurophysiological events and states that cause actions—in which case reason explanation (and causation) seem superfluous, or they are not identical to them—but this has the consequence that every action is causal both by a reason and by the corresponding neurophysiological events and states. Most philosophers reject the possibility of this sort of causal overdetermination. But then, any sense in which reasons causally explain actions becomes obscure. This conclusion has been expressed differently (e.g., as the claim that the mental is causally irrelevant or epiphenomenal or that mental properties have no causal role or that they are inert).

Another reason for dissatisfaction with the causal theory is a sense that, on examination, it appears to leave the agent out of the picture. The causal theory, this objection goes, makes the agent appear, at most, as the locus for the causal transactions between events (or states) that constitute her reasons and her actions. Thus, on this picture, the agent herself turns out to be passive where her actions are concerned! This would be an ironic consequence for the causal theory to have since it was introduced supposedly to explain what it is for someone to act for a reason. Thus, unless it can bring the agent back into the picture, the causal theory does not explain why agents should be thought to act at all, let alone to act freely or with moral responsibility. This objection has led some to revisit the traditional concept of agent causation—that is, the view that there is an irreducible relation between an agent and her actions, while others have sought to revise this concept to free it from its traditional problems. But most contemporary philosophers tend to be skeptical of the very idea of agent causation, because, they claim, agent causation is not an explanatory notion and it is reducible to event causation: To say that an agent causes an event, they would claim, is just to say that an action of the agent (an event) does. Part of what is at issue in these problems is how to specify the conditions under which an agent can be said to have
performed a *free action*—an action for which he can be held morally responsible. This brings us to the much-debated old issue of the compatibility or incompatibility of free will and moral responsibility with determinism.

**Agency, Freedom, and Moral Responsibility**

The traditional debate on these issues assumed that moral responsibility requires freedom and that freedom requires *alternative possibilities*—the ability to act otherwise. The question then was whether, if determinism is true, freedom and hence moral responsibility are possible and justified, respectively. That is, if actions have causes and if all causes necessitate their effects, then actions seem to be determined and agents, it would seem, are not free to act (or to refrain from acting) and hence should not be held to be morally responsible for their behavior. (The causal theory would face this problem because it says explicitly that actions are events caused by reasons, but anyone who accepts that the events that agents are said to cause could be necessitated faces similar difficulties.)

The abundant recent literature on the topic has a distinctive flavor for several related reasons. One is that in the mid-1960s, Harry Frankfurt published a paper challenging the idea that moral responsibility does indeed require the ability to act otherwise. In that paper, Frankfurt claims to offer an example where people on both sides of the debate about the compatibility of free will and determinism would intuitively agree that the agent *is* morally responsible for his action even though he *could not* have done otherwise. Much ink has been spilled on whether Frankfurt’s challenge succeeds, whether it begs the question against incompatibilists (who hold that free will and determinism are incompatible), and indeed whether the thought experiments on which his arguments depend are cogent. Be that as it may, this has led to a revision of the concept of an agent (or a person). Frankfurt, for example, argues that agents (or persons) are characterized by their capacity for second-order mental states (that is, attitudes one has toward first-order attitudes, such as my belief that I want to be liked by all or my desire not to act according to that desire), and he has developed an account of agency in terms of *identification*. In this view, an action is properly an agent’s only if the latter identifies with the “springs” or sources of his or her action—that is, has second-order attitudes (roughly, of endorsement) toward the first order attitudes of desire, which are said to be the origins of his actions. Since then, many philosophers have tried to develop accounts along these lines, which are characterized by the prominence they give to the concepts of control and autonomy.

A related feature of current debates on free will is that they are often informed by the thought that free action must be, in some substantial sense, action that comes about as a result of agents’ *responsiveness* to reasons: That is, agents must be acting for the reasons they take themselves to have. This of course raises the issue what *reasons* for action are.

**Actions and Reasons for Acting**

A recent development in the philosophy of action can be traced to work on practical reasoning (reasoning about how to act) and normativity (a much used but underdefined term that concerns, roughly, the kinds of requirement that norms, reasons, and/or values place on agents). This has resulted in a reexamination of the previously prevalent conception of reasons for action. Since the 1960s, most philosophers of action conceived of an agent’s reason for acting, called the agent’s *motivating* (or explanatory) reason, as a combination of a belief and a desire. Beliefs and desires were usually thought of as *propositional attitudes*: mental states of agents that consist in the agent’s taking a certain attitude to a proposition (to the *content* of the attitude). Thus, a motivating reason was typically construed as a mental state with a content, which could (perhaps via a triggering event) cause events (actions, or intentions that in turn cause actions). But this conception of reasons for action has come under pressure in recent years.

This pressure has largely come from considerations about the relation between motivating reasons (the reasons for which agents *actually* act) and normative reasons (the reasons there are for agents to act in certain ways, regardless of whether they recognize these reasons)—considerations that seem to support the conclusion that motivating reasons cannot be mental states or indeed psychological entities of any type. The arguments for this conclusion vary but its defenders tend to agree that what motivates an agent is not *that be believed* something but rather *what be believed*. We are motivated, say, by what we believe (e.g., that the food is poisoned) and not by
our believing it. And what we believe (i.e., that the food is poisoned) is not a mental state. To put the point differently, the reasons that motivate agents are not their own mental states but rather aspects of reality (as some would say, not mental states but facts). And while, traditionally, the response might have been that this rests on a confusion between normative and motivating reasons, the insight that normative and motivating reasons must be closer in character than they have traditionally been supposed to be undermines this response.

Thus, considerations about normativity have put pressure on the traditional view of motivating reasons as mental states with causal powers. However, it should be noted that the attempt to bring together normative and explanatory reasons calls for an explanation of the connection between motivation by reasons, which are not mental states, and psychological explanations that seem to refer to mental states. In short, reflection about normative and motivating reasons suggests that we need to think afresh the connection between deliberation, normativity, motivation, and reason explanation.

This brief overview of the philosophy of action leaves out many important issues, such as the problem of akrasia, the role of emotion and the unconscious in human action, the topics of collective or social action, rationality and rational choice, and the rise of so-called experimental theory of action. But both what is included and what is left out show how rich, interesting, and relevant this field of philosophy is.

Maria Alvarez

See also Action and Bodily Movement; Collective Action; Explanation of Action; Freedom of Action; Mental Action; Phenomenology of Action

Further Readings


Physicalism

Physicalism is the claim that everything in the universe is physical. That doesn’t mean that there is nothing living or mental; otherwise, physicalism would attract few believers. Instead, physicalists hold that the living and the mental are types of physical things. Physicalism is viewed as the default assumption for scientific approaches to the understanding of mental life. The first main section of this entry focuses on what makes something physical and refines our understanding of the claim that everything is physical. The second focuses on the main argument for this position, while the third considers the main problem with physicalism.

The Nature of Physicalism

Earlier physicalists—such as Thomas Hobbes—provided a substantive characterization of the nature of the physical as, for example, occupying space and/or possessing mass. With the development of physics, the emphasis changed. The objects and properties characterized by modern physics seemed very different from our everyday understanding of physical objects, yet it seemed a mistake to take developments in modern physics to prove that physicalism was incorrect. Thus modern day physicalists, such as J. J. C. Smart, sought to define physicalism in terms of physics. They recognized that physics may conceivably develop in ways that would recognize nonphysical entities and that not everything physical is identified by physics.

For example, the properties of being a mammal or of being a chair seem to be physical but aren’t identified by physics. So they eventually came up with a characterization on the following lines: A property is a broadly physical property if and only if either it is identified by a physics that sufficiently resembles our own current physics (where by this they mean the postulates of that body of physical theory broadly accepted in the late 20th and early 21st century, including the special and general theory of relativity and quantum mechanics) or it supervenes on those properties identified by physics (supervenience will be described below). The basic idea is that a physics that postulated nonphysical entities would be a significant departure from, and hence not resemble, current physics as just understood. Physical objects and events are those that only have physical properties in the specified sense. Narrowly physical properties are just those identified by physics of the type indicated.

Great energy was then devoted to identifying what kind of supervenience was involved. Supervenience is a type of covariation between members of specified families of properties, for example, between properties of arrangements of bricks and properties of being brick constructions such as walls. Although identity between properties is a limiting case of supervenience, supervenience does not require identity between them; otherwise, it would be ill-suited to the task at hand since it would require that all properties are identified by physics. Nor can this kind of supervenience be simply a lawful relationship between properties. Those who deny that physicalism is true don’t have to deny that there is such a relationship between physical and nonphysical properties. For example, suppose you believed in ghosts and that, in the presence of ghosts, there would be a drop in temperature. Then there would be a lawful relationship between ghostly properties and physical ones but this, alone, would not make ghosts physical. Thus, many philosophers appealed to a metaphysically necessary relationship, for example, the kind that holds between having angles adding up to 180° and being three sided. No matter how things might otherwise might be, the thought ran, if two universes were identical in their arrangement of the properties identified by physics (and no other properties were added), then the distribution of all other properties, including, specifically, mental properties, would be fixed.

However, although this kind of supervenience labeled an intimate connection between types of properties, the nature of the relationship remained mysterious. When there is identity between those properties identified by physics and properties identified by other means, then no further explanation is needed. When there is no identity, the need for some kind of explanatory connection is far more pressing. Talk of the properties identified by physics constituting all the other properties provided one kind of popular answer—just as the arrangement of pieces of cardboard can constitute a box or the arrangements of bricks make up a wall. The desire for an explanatory connection, and attendant talk of constitution, was popular for another reason. Many felt that physicalism must involve the idea that the properties identified by physics determined the presence of all the other properties and were, in some sense, the fundamental properties of the universe. As a way of cashing this out, the laws identified by physics were taken to be the fundamental laws of the universe.

While these additional features (the presence of an explanatory connection, constitution, physical laws being fundamental) certainly capture the kind of physicalism that many find attractive, it is doubtful whether they should be taken to be definitive of physicalism for, at least, two reasons. The first is that this type of characterization of physicalism rests on a highly ordered conception of reality in which everything ultimately is based on physics. Physics may be important in characterizing the nature of the physical but it is by no means clear that reality does have such an ordered character. Instead, physics could be seen as one of a range of sciences each with its own subject matter and various kinds of relations between the subject matters. If physics was just the science of the very small, then it is easy to see how the universe might be thought to be based on physics, but the subject matter of physics is broader than that with its own emergent phenomena—such as quantum superposition and entanglement—reaching across wide areas of space. The second reason for being dubious is that this type of characterization of physicalism ignores the possibility of emergent physicalism. Emergent physicalism rests on the possibility that other sciences—psychology, biology, chemistry—may identify emergent causal powers of arrangements of narrowly physical properties that cannot be understood in terms of the kind of
arrangements of properties that have been of interest to physicists.

The Main Argument for Physicalism

The main arguments for physicalism have derived from causality. Often, it has been put, especially in the past, as a worry about whether something nonphysical—ethereal, or outside space and time, or the like—could influence the passage of physical events. So formulated, it has little real bite as there is no reason to expect that fundamental causal relations should be readily intelligible and so no reason to deny that the nonphysical could influence the physical. More recently, systematic causal considerations have been advanced that trade on no such assumption. These are laid out in what is generally dubbed the overdetermination argument with its focus, invariably, on mental properties that are taken to be the main candidates for being nonphysical.

Consider the feeling of awful pain you have (P) after having just put your hand in a flame and, putting it neutrally, the neural state in the brain that neuroscientists have identified as concerned with pain (N). Suppose that P is not identical to N (or anything else physical) but is correlated, perhaps, lawfully with it. The supposition that P is not identical to N might be tempting because it can seem that, while neuroscience can explain why we feel pain, its descriptions cannot exhaust how pain feels to us. As a result of being in pain, you withdraw your hand from the flame. What's the cause of this withdrawal? If you say P or N + P, then you must accept that there is something identified by physics (some component[s] of the arm) that is partly caused by something nonphysical: P. But then, either this conflicts with the apparently plausible thesis that the world described by physics is causally complete—only those things identified by physics are required in the causal explanation of other things identified by physics—or we have overdetermination (which promises to be systematic and extensive because the argument can be run for many mental properties and their neural associates), or P is epiphenomenal. We would have overdetermination if both P and N individually were sufficient to explain the activity in the arm—just as somebody might be killed by two assassins’ bullets each of which, individually, would have been enough to kill him or her. P would be epiphenomenal if it had no effect on the arm, in roughly the way that the color of a piece of fruit has no effect on how much it weighs.

The unattractiveness of these consequences has been questioned. Perhaps the causal influence of P is hidden or revealed in systematic ways that need not conflict with the apparent causal completeness of physics. Maybe systematic and extensive overdetermination is not so bad because it can be explained by psychophysical laws. It may not be a central component of our understanding of our mental lives that P has an influence. Instead, epiphenomenalism may be true.

There has also been substantial debate over whether those physicalists who deny that all properties are identical to those identified by physics are in a better position to avoid the argument. After all, they too allow that P is not identical to N and P is a cause of the behavior. Responses to this latter worry center around taking causation between events not identified by physics as constituted in some way from causation between events that are identified by physics, in much the same way that broadly physical properties are taken to be constituted from narrowly physical properties.

Whatever the success of the argument from overdetermination and its ramifications for physicalism, physicalism remains an independently attractive position because it is seen as acknowledging the importance of scientific investigation into the nature of the mental and other phenomena and the central importance that physics can play in such investigation.

Problems for Physicalism

Although challenges to physicalism have been derived from the nature of life, freewill, and intentionality (the power of thoughts to represent the world), the main challenge is generally agreed to be consciousness and phenomenal consciousness—the what it’s likeness of consciousness—in particular. Physical descriptions of the character of phenomenal consciousness have been felt to be inadequate by many philosophers. The point has been put in various ways, for instance, being unable to imagine what an alien creature’s experience is like (such as a bat) even if we have a full physical description of it or being unable to know what it is like to experience a red tomato if color blind. Each of these ways involves additional factors such as the limits of imagination or the nature of knowledge, which can obscure the point. The clearest way to present the challenge is to observe that there seems to be an explanatory gap between the nature of the physical properties of the brain (and its surroundings) and the nature
of color experience (say). We can’t understand why color experience supervenes on these other properties whereas, in contrast, we can understand why liquidity supervenes on the relatively weak forces between layers of molecules.

The obvious explanation of an explanatory gap is that, in fact, we are talking of two wholly distinct types of properties: physical properties and the properties of experience. This is an answer that physicalists cannot give. So their basic response has been to argue that it only appears to us that there are radically different properties because the concepts we have concerning the nature of our own experiences—often dubbed phenomenal concepts—have distinctive features not shared by concepts of either the entities identified by science or everyday objects in the world. Nevertheless, it has proved to be very difficult to identify the precise features of phenomenal concepts that explain why there seems to be an explanatory gap.

According to one line of thought, phenomenal concepts share with demonstrative and indexical concepts (e.g., that, I) the feature of being applicable without appeal to description. With an indexical, we don’t need to describe ourselves in any particular way in order to refer to ourselves by the pronoun I. Likewise, with a demonstrative, we don’t need to, indeed find it hard to, describe our experience in any way in recognizing it as feeling like that. Unfortunately, while “I am Paul Noordhof” and “A-delta firing [one type of pain] feels like that” may both be informative, the latter remains additionally puzzling. Why should it feel like that? Lack of a satisfactory response has seen physicalism moving from being apparently irresistible to being a default hypothesis that remains to be fully justified.

Paul Noordhof

See also Access Consciousness; Emergence; Explanatory Gap; Mind-Body Problem; Reductive Physicalism

Further Readings


Placebo Effect

The placebo effect is a change produced by the administration of a substance or physical procedure that is not produced by the physical properties of that substance or procedure. It is, instead, an effect produced by the psychological effects of administering the treatment. A placebo is a substance or procedure that cannot produce a particular effect by virtue of its physical properties. This entry describes the history of the placebo concept, its use in clinical trials, factors affecting the magnitude of the placebo effect, and theoretical accounts of how placebo effects are produced.

The word placebo comes from the Latin word placeo meaning “I shall please,” and for centuries, it was assumed that placebos could placate troubled patients but not produce any real changes. It was not until the mid 20th century that researchers began to appreciate that placebos might produce changes in symptoms. Once that realization became widespread, placebos became commonplace in the process of approving new drugs. Randomized controlled trials (RCTs) soon became the gold standard for testing new medications.

In an RCT, subjects are randomly assigned to at least two groups, an active treatment group and a placebo group. Some trials may include three or more groups. For example, two different drugs may be compared to each other and to a placebo, or subjects might be randomized to receive different doses of the same drug. Assignment to treatment condition is double-blind, which means that neither the physician nor the patient is told to which group the patient has been assigned. The placebo is made so as to be indistinguishable from the real drug. It is the same color, size, and shape. The idea is to keep
all psychological variables constant between the active and placebo treatment conditions so that any differences in the effects obtained can be unequivocally ascribed to the active physical properties of the treatment.

Usually, placebos are physically inert, but active placebos are sometimes used. An active placebo is an active substance that can produce side effects but that should have no physical effect on the condition being treated. The purpose of using an active placebo is to prevent patients from breaking blind. Breaking blind occurs when patients are able to figure out which group they have been assigned to, perhaps because of the side effects that the real drug produces.

Magnitude of the Placebo Effect
In 1955, Henry Beecher, a pioneer researcher of the placebo effect, wrote an article titled “The Powerful Placebo,” in which he asserted that one third of all patients respond to placebos. In fact, this turns out to be wrong. The strength of the placebo effect depends on a number of factors. Depending on these factors, a placebo might have no effect at all, or it might affect 100% of participants. Most important, the strength of the placebo effect depends on the condition being treated. For example, placebos can duplicate more than 80% of the effect of antidepressant drugs and 50% of the effect of painkillers, but they do not seem to have any effect at all on blood sugar levels in the treatment of diabetes. In general, placebo effects are more likely to occur in responses that are consciously experienced than in those that are not. Other factors affecting the placebo response include the color of the placebo, its price and name, the apparent dose, the strength of the drug for which the placebo is a substitute, and the mode of administration.

Color
The color of a placebo can influence its effects. When administered without information about whether they are stimulants or tranquilizers, blue placebo pills produce tranquilizing effects, whereas red placebos induce stimulant effects. Patients report falling asleep more quickly after taking a blue capsule than after taking an orange capsule, and red placebos seem to be more effective pain relievers than white, blue, or green placebos.

Price and Name
Placebos with a recognizable brand name are more effective than placebos described as generic drugs, and more expensive placebos are more effective than cheaper ones. Perhaps this is why pharmaceutical companies can successfully market their brand-name products even when less expensive, generic equivalents are available.

Dose
There are two ways in which dose can make a difference in the effects of placebo treatment. One is the number of placebos given. In one study, placebos prescribed for the treatment of ulcers were more effective when the patients were asked to take four pills per day than when patients were asked to take two pills per day. Another way in which dose can make a difference is in the degree to which factors deemed to be responsible for the placebo effect are involved in its administration. For example, the placebo effect is thought to be dependent on the presence of a supportive therapeutic relationship with the health care provider and on the confidence that the provider communicates about the effectiveness of treatment. One study showed that placebo treatment of irritable bowel syndrome was substantially more effective when the initial interview with the patient was longer and when the clinician was warm and supportive and expressed confidence in the treatment.

Strength of the Active Drug
A placebo given in the guise of a more potent drug is more effective than one given in the guise of a less potent drug. Recall that placebo painkillers are about half as effective as real analgesic medication. What that means is that placebo aspirin is half as effective as real aspirin, and placebo morphine is half as effective as real morphine. Since morphine is more potent than aspirin, placebo morphine is more effective than placebo aspirin.

Mode of Administration
Placebo injections are more effective than placebo pills, and sham surgery is more effective than either injections or pills. Placebo surgery involves cutting patients open and then sewing them up again but not performing the surgical intervention. Sham operations have been used as control procedures
in clinical trials evaluating the effectiveness of mammary ligation in the treatment of angina and arthroscopic surgery in the treatment of osteoarthritis of the knee. In both of these applications, the sham surgery was found to be as effective as the real surgery. Mammary ligation is no longer used as a treatment for angina. Although arthroscopic surgery for osteoarthritis of the knee is still performed, it has become controversial because of the research showing placebo surgery to be equally effective.

**Individual Differences**

In any given clinical trial, some people will respond to a placebo and some will not. This finding has led researchers to search for the characteristics of placebo responders. The question is, how do placebo responders differ from nonresponders? Many believe that if this question could be answered, placebo responders could be screened out from clinical trials, thereby making it easier to detect the effects of real drugs.

The search for personality characteristics of placebo responders has been largely unsuccessful. Most studies have failed to find significant differences in the traits that were measured. Exceptions include studies indicating that placebo responsiveness might be linked to an acquiescent response set (i.e., a tendency to say yes) and to dispositional optimism. Optimists seem more likely to experience beneficial effects from placebos, whereas pessimists are more likely to experience negative effects. This latter effect is often called the nocebo effect. The nocebo effect occurs when negative expectancies produce unwanted consequences. Evidence supporting the classical conditioning model of placebo effects comes from studies in which a placebo cream was used to lower the feelings of pain produced by an experimental pain stimulus. Conditioning is accomplished by surreptitiously lowering the intensity of the pain stimulus when it is applied to the part of the body that to which the placebo cream has been applied. Repeated conditioning trials of this sort result in an enhanced placebo effect when the pain stimulus is later administered at full intensity.

**Theories of Placebo Effects**

There are two main psychological theories of placebo effects: classical conditioning and response expectancy.

**Classical Conditioning**

Classical conditioning was discovered by the Russian physiologist Ivan Pavlov at the end of the 19th century. The most famous example of classical conditioning is that in which the presentation of food to a dog is paired repeatedly with the sound of a bell. After a number of such pairings, the bell acquires the ability to cause the dog to salivate even when food is not presented. In this example, the food is termed an unconditional stimulus (US), salivation in response to the food is termed an unconditional response (UR), the bell is a conditional stimulus (CS), and salivation in response to the bell is called a conditional response (CR).

Classical conditioning was discovered by the Russian physiologist Ivan Pavlov at the end of the 19th century. The most famous example of classical conditioning is that in which the presentation of food to a dog is paired repeatedly with the sound of a bell. After a number of such pairings, the bell acquires the ability to cause the dog to salivate even when food is not presented. In this example, the food is termed an unconditional stimulus (US), salivation in response to the food is termed an unconditional response (UR), the bell is a conditional stimulus (CS), and salivation in response to the bell is called a conditional response (CR).

The classical conditioning model of placebo effects is based on the observation that active treatments are always administered in some type of vehicle (e.g., a pill, capsule, or injection). Thus, the active treatment (the US) is paired with the vehicle in which it is administered (the CS). As a result of these pairings, the effect of the treatment (the UR) comes to be administered by the vehicle alone as a conditional response. Evidence supporting the classical conditioning model of placebo effects comes from studies in which a placebo cream was used to lower the feelings of pain produced by an experimental pain stimulus. Conditioning is accomplished by surreptitiously lowering the intensity of the pain stimulus when it is applied to the part of the body that to which the placebo cream has been applied. Repeated conditioning trials of this sort result in an enhanced placebo effect when the pain stimulus is later administered at full intensity.

**Response Expectancy**

Response expectancies are anticipations of subjective experiences and other automatic, unintentional responses. A wealth of research indicates that response expectancies tend to produce the expected responses in the form of self-fulfilling prophecies. For example, the expectancy that one will experience a panic attack seems capable of inducing one, and expected pain reduction leads to experienced pain reduction.
According to expectancy theory, placebos produce placebo effects by altering response expectancies. For example, placebo analgesics produce expectancies of reduced pain, placebo antidepressants lead one to expect to feel less depressed, and placebo caffeine produces expectancies of feeling aroused. These response expectancies then elicit the expected responses in the form of placebo effects. Expectancy theory seems to do a particularly good job of explaining some of the factors that affect the strength of the placebo effect. People expect brand name medications to be more effective than generic equivalents, for example, and they know that morphine is more potent than aspirin.

**Conditioning and Expectancy**

Although conditioning theory and expectancy theory are sometimes pitted against each other as alternatives, modern forms of conditioning theory are compatible with response expectancy theory. According to these formulations, conditioning is not an automatic, unconscious process. Instead, it works by producing expectancies, which in turn elicit conditional responses. In the example of Pavlov’s dogs, conditioning trials lead the animal to anticipate food whenever the bell is rung. It is the anticipation of the food that then produces the salivation.

As applied to placebo effects, the integration of conditioning theory and expectancy theory holds that conditioning trials are one means—perhaps the most effective—of producing the expectancies that then elicit placebo effects. Consistent with this approach, studies have shown that conditioning procedures aimed at enhancing placebo pain reduction produce expectancies for reduced pain and that the expectancies are correlated with the amount of pain subsequently reported.

**Future Directions**

The importance of the placebo effect is now widely accepted. Besides its use as a means of distinguishing between drug effects and psychological effects in clinical trials, the placebo effect is increasingly recognized as a component of active treatments. For that reason, the task of harnessing the placebo effect so that it can be used clinically is important. Hindering this effort is the ethical problem of deceiving patients and the general assumption that deception may be necessary for effective use of the placebo effect in clinical practice. This has led to an ongoing debate about whether deception might be justified when it is likely to benefit the patients who are being deceived.

*Irving Kirsch*

**See also** Behaviorism; Reinforcement Learning, Psychological Perspectives

**Further Readings**


Planning in Language Production

The planning and articulation of speech often overlap in time, much as the downloading and playing of video content overlap when streamed over the Internet. Specifically, while speakers articulate the initial words of an utterance, they plan its subsequent parts. Research has focused on the production of novel, isolated picture descriptions such as “green cat” and “a turtle is squirting a mouse.” Little work has addressed the planning of utterances for purposes other than description, narration, or sentence completion, such as utterances with primarily social functions and those containing conventional phrases, such as “How are you?” This entry summarizes the research on the time course of utterance planning for adult, unimpaired, proficient, native speakers of (primarily) Indo-European languages.

Scope

Speakers appear capable of planning and storing in memory (i.e., buffering) a representation containing the sounds of an entire utterance prior to articulating it, but for multiword utterances they do not do so spontaneously. Just as pauses in playback and slow connection warnings indicate that downloading has not stayed ahead of playback, the prevalence of disfluencies (e.g., silent pauses and delay signals such as um and uh) suggests a limited scope of planning and buffering. Moreover, buffering material is resource demanding and feedback from listeners may often make buffered material irrelevant, similar to downloading a whole movie that a viewer cancels after a few minutes.

Message

The preverbal message for a descriptive utterance contains a topic (what an utterance is about) and information that the speaker wishes to express about the topic. As the topic is usually the first thing mentioned, its message representation is probably fully specified and the words and sounds used to express it retrieved before any part of the utterance is articulated. Suggesting that speakers outline a message prior to utterance onset, initial pitch varies with sentence length, while sentence length is largely a function of message content and complexity. Also, having outlined preverbal messages allows speakers to order simple content before complicated content in languages structured like English (e.g., “The philanthropist donated to a charity a wide array of products and services that would prove to be very useful”) and complex content before simple in languages such as Japanese. Patterns of spontaneous word anticipations, exchanges, and subject-verb agreement errors suggest that speakers rarely plan or simultaneously consider messages beyond the clause that they are currently articulating.

Speakers need to simultaneously represent and maintain message representations that correspond to entire noun phrases to fluently produce prenominal modifiers (e.g., the long, brown hair) and conjoined noun phrases (dog and cup). Speakers can modify the content of a noun phrase while articulating its initial parts, but this leads to disfluencies when there is insufficient time to encode the new content. For example, a speaker may fluently utter, “the alien with the small spots,” and only decide to mention the size of the spots around the time he or she starts articulating alien. However, if the speaker wants to say “the small alien” fluently, he or she needs to consider size at least a half second before the onset of the.

Structure and Order of Mention

Message representations are not thought to contain any intrinsic order but primarily relationships between concepts that differ in availability. Some theories posit that bits of grammatical structure are accessed via selection of content words (e.g., nouns, verbs) and then combined. Other theories see the structure of an utterance as emerging from a sequence of decisions about what part of a message to put into words next, constrained by what the speaker has already planned and the language spoken (e.g., in English a direct object follows its verb rather than precedes it).

Words

In spontaneous speech, speakers often hesitate before articulating words that have many near synonyms to choose among. Even in fluent speech, experiments suggest that nouns are selected shortly before or during articulation of their encompassing noun phrase.
However, it is less clear when speakers select verbs, but speakers need not select them before encoding other parts of an utterance. Disagreement centers on whether speakers can select words from the same grammatical class within a phrase (i.e., two adjectives or two nouns) simultaneously or if they select them one at a time with selection affected by such composite message representations.

Closed-class, functional elements such as articles \((a, the)\) and verb tense markers \((-ed, -ing)\) appear to be selected via a different process than content words are. The extremely high frequency of use for functional elements should make their retrieval very fast and accurate. Indeed, they appear to take relatively little time to plan. However, these elements are acquired later than content words in language acquisition and may be selectively impaired with brain damage. Perhaps contributing to their difficulty in acquisition and planning is the abstractness of their meanings (e.g., I walked vs. I was walking). Theories of language production tend to link the retrieval of functional elements to the same mechanism that determines grammatical structure.

The selection of a word and retrieval of its sounds appear separable although they nearly always occur in immediate succession. For example, speech errors often involve selecting an unintended semantically related word but retrieving its sounds perfectly (e.g., substituting bike for car) or selecting the intended word but then flubbing its sounds (saying urvan for urban). Sometimes speakers are able to select a word successfully but then fail to retrieve most of its sounds, resulting in a feeling of having a word on the tip of the tongue.

**Sounds and Movement**

Theories often distinguish between retrieving the individual sounds of a word, organizing them into syllables or other rhythmic patterns, and computing the movements needed to articulate them. Sound-related processing appears to have a scope of about one second or up to two content words. Brain-damaged patients who have severe deficits in short-term memory for phonological information nonetheless tend to converse normally, suggesting that there is little need to retrieve sounds more than one message element or phrase in advance.

Current issues in planning concern the role of working memory beyond maintaining messages or phonological plans, the generalizability of results from simple descriptions of objects, variation in planning scope with speech rate, effects in articulation durations, similarities with other forms of action planning, and developing theories that are less dependent on discrete symbolic representations.

Zenzi M. Griffin

**See also** Aphasia; Language Production, Incremental Processing in; Production of Language; Prosody in Production

**Further Readings**


**Political Psychology**

In its broadest sense, political psychology addresses how human nature shapes political life. Philosophers have been investigating this question for a very long time, dating at least to the ancient Greeks: Plato’s *Republic* and Aristotle’s *Politics* are both concerned with the limitations of human nature and the implications for the design of political systems. Later, European philosophers such as Thomas Hobbes and Jean-Jacques Rousseau differed sharply in their assumptions about human nature and, hence, in the political conclusions they drew (with Hobbes taking
the view that people’s inherent tendencies to aggress required a strong state and Rousseau contending that people are naturally inclined toward peaceful coexistence). This entry describes modern political psychology, beginning with Theodor Adorno and colleagues’ *The Authoritarian Personality*. It then covers critiques of Adorno’s approach, the “rediscovery” of ideology in the 1980s, and modern research on ideology and personality as well as on incidental factors that affect policy preferences and voting behavior. Finally, charges of ideological bias among political psychologists are discussed.

**Ideology and Personality**

Political psychology in its current sense emerged as a discipline following World War II. Shocked by the ease with which prewar Germany had turned to authoritarianism, researchers sought to understand the personality factors that predisposed individuals to support authoritarian leaders and regimes. Generally, they adopted a Freudian model that regarded early childhood as determinative of adult personality. The most influential work in this tradition was Adorno and colleagues’ *The Authoritarian Personality*. Published in 1950, it was an ambitious attempt to link personality characteristics (and the childhood experiences that were thought to have produced these characteristics) with political ideology, and it exerted a singular influence on the nascent field of political psychology. Adorno and colleagues were interested in the psychological underpinnings of anti-Semitism and, more broadly, in support for authoritarian regimes. As such, their research focused mainly on the personality characteristics of authoritarians, specifically authoritarian submissives—the followers rather than the leaders in authoritarian regimes. A particular concern was the emergence of authoritarianism in a democracy, and so Adorno and colleagues studied U.S. citizens rather than citizens of actual authoritarian regimes. The ultimate goal was to understand the personalities of supporters of authoritarianism and thus to understand how authoritarian regimes can arise and rapidly gain support, as was the case in prewar Weimar Germany.

*The Authoritarian Personality* drew on questionnaire measures (the most well known is the potentiality for fascism, or F, scale), semi-structured interviews, and projective tests such as the Thematic Apperception Test (TAT). Adorno and colleagues argued that these diverse methods revealed a coherent portrait of potential authoritarians: Unrealistically positive about themselves and about parental figures, they redirect hostility toward socially marginalized outgroups. They value status and material success. Mentally, they are rigid, inflexible, and intolerant of ambiguity, preferring to see the world in black and white.

In the wake of *The Authoritarian Personality*, there was a flurry of interest in personality correlates of ideology and in the validity of the specific methods used by Adorno and his team. Although there were numerous criticisms of Adorno and colleagues’ methods, especially of the reliance on subjectively scored tests and interviews, the most lasting damage was inflicted by political psychology’s increasing skepticism about whether people (or, at any rate, U.S. citizens) possessed coherent ideologies to begin with. The most damning critique, advanced by Philip Converse in “The Nature of Belief Systems in Mass Publics,” was that most people showed little or no evidence of a coherent ideology. With the exception of the roughly 10% of the population who were, in Converse’s terminology, ideologues or near ideologues, survey respondents showed little evidence of a coherent ideological belief system—just because people took a liberal (or conservative) stance on an issue did not mean that they would take similar positions on related issues. Furthermore, the temporal stability of attitudes was quite low: If a respondent was for, say, affirmative action in 1974, odds were little better than chance that he or she would still hold that position in 1976. Even awareness of ideological terms was strikingly low—very few respondents could correctly assign the terms liberal and conservative to the Democratic and Republican parties and explain what each meant. Converse’s argument was extremely influential, and until the early 1980s, researchers made few serious attempts to study ideology. Rather, survey researchers of the 1960s and 1970s attempted to predict political attitudes and voting behavior on the basis of other factors, most notably demographics and party identification.

**The Rediscovery of Ideology**

Possibly because of increasingly sharp partisan divides in the United States, the 1980s saw a resurgence of interest in ideology, especially conservative
ideology. Instruments such as Robert Altemeyer’s right-wing authoritarianism (RWA) scale, James Sidanius and colleagues’ social dominance orientation (SDO) scale, and John Jost and Erik Thompson’s economic system justification scale were intended to measure broad attitudes toward authoritarian government and social and economic inequality.

Recently, John Jost and his colleagues have argued that conservative ideology—which they define as a resistance to change and a tolerance of inequality between social groups—is motivated by the psychological need to control uncertainty and threat. In its conception of conservatism as a product of deep psychological needs, this model is akin to Adorno and colleagues’ view of support for authoritarianism as the result of people’s refusal to acknowledge negative feelings toward the self, parents, and authority figures. However, there are important differences: Adorno and colleagues took a Freudian perspective that treated the authoritarian personality as the product of an overly punitive parenting style. In contrast, Jost and colleagues see conservatism as a way of imposing order on an unpredictable, dangerous world. In “Political Conservatism as Motivated Social Cognition,” they summarize a great deal of research in the last 50 years showing a relationship between conservatism and personality constructs relevant to needs for safety and order: Conservatism correlates positively with measures of anxiety, positively with aversion to ambiguity and uncertainty, and negatively with sensation seeking and openness to new experiences. Although Jost and colleagues’ conclusions (and Jost’s focus on conservatism in particular) have been vigorously debated, it seems clear that the study of ideology and the psychological motives that underlie it are once again central to political psychology.

This is not to say that political psychologists see people as having completely coherent political views. In fact, a robust stream of research has examined the effect of normatively irrelevant factors on people’s policy evaluations and voting behavior. For example, Geoffrey Cohen has shown that people often attend to the political party advocating a policy, rather than to the policy’s actual content, when deciding whether to support or oppose it. Even more startling is the effect of completely irrelevant contextual factors on people’s actual voting behavior. Candidates who were listed first on the ballot gained an average advantage of 2.5 percentage points across 118 Ohio races, and Arizonans who voted in a school (as opposed to those whose polling place was in another location) were more likely (by about 2 percentage points) to support an increase in the sales tax to increase school funding.

Is Political Psychology Biased?

It is probably fair to say that most political psychologists are personally politically liberal. Throughout the history of modern political psychology, this has led critics to contend that the field is prone to political bias. One kind of critique has focused on how political psychologists formulate research questions. The Authoritarian Personality was criticized for focusing on political conservatives as an other to be explained; similarly, Jost and colleagues’ view of conservatism as motivated cognition has been criticized for assuming that the political views of conservatives, not of liberals, demand explanation in terms of underlying psychological needs and motives. A second type of critique deals with how key concepts are defined—for example, Philip Tetlock has argued that the concept of modern racism (as advanced by David Sears and colleagues) is inherently politically biased in that it labels people who are opposed to income redistribution and school busing as modern racists by definition. The validity of these critiques has been hotly debated, but recent work in political psychology has been influenced by them, at least implicitly. One recent example is the work of Jon Haidt and colleagues on the moral foundations underlying the political views of liberals and conservatives, which argues that the two groups rely on divergent but equally valid moral intuitions.

Conclusion

Political psychology—the study of how human nature shapes political life—has attracted the interest of philosophers dating back to the ancient Greeks. More recently, empirical researchers have sought to apply the tools and paradigms of modern psychology to the study of political beliefs and behavior. Political psychology emerged as a discipline following World War II with Theodor Adorno and colleagues’ studies of the “authoritarian personality.” Although interest in ideology waned for a time as researchers focused on nonideological predictors of voting behavior, personality-driven research into political ideology, especially conservative ideology, has enjoyed a recent revival. This focus on conservative ideology has been criticized as ideologically
Preconscious Free Will

This entry presents scientific evidence that the conscious feelings of wishing and deciding that we usually think of as determining our voluntary acts are not themselves responsible for those acts, although they may roughly but accurately represent preconscious mental processes that are responsible for such acts.

How Free Will Is Normally Experienced

Within philosophy, free will is commonly thought of as “free” to the extent that its operations are not entirely determined by physical causal laws. However, freely willed actions are not usually experienced as being entirely free of determining factors. Rather, choices, decisions, and actions are normally experienced as operating, with some degrees of freedom, within complex mental, physical, and social constraints. Consequently, psychologists have typically focused their interest on the systems that enable humans to have the freedom to choose, decide, and act that they actually experience themselves to have—a form of constrained free will. Studies of voluntary action in humans have made it clear that such systems need to include inner needs and goals, a global knowledge store (based on previous interactions with the world), processes for modeling current inner and external states of affairs, alternative strategies for action, methods for assessing the likely success of alternative strategies in the light of existing physical and social constraints, and the ability to learn from experience. Although such systems follow deterministic principles, their operation can be partly self-organizing and flexible, and their complexity can allow sufficient degrees of freedom to accurately model the ability to make choices and decisions, within the available alternatives, that humans actually experience. Given this, there is nothing within current psychological understanding of the mind, viewed as a complex system, which rules out a form of constrained free will—a position known in philosophy of mind as compatibilism.

Distinguishing Free Will From Conscious Free Will

Free will does, however, have to be distinguished from conscious free will, because in principle, the operations of such decision-making systems do not have to be conscious. If the detailed information processing involved could be sufficiently well specified, it could, for example, operate equally well in a nonconscious robot. Neuropsychological findings have also cast doubt on the role of the conscious experiences associated with willing and deciding in the operation of the volitional processes themselves.
It has been known for over 40 years that voluntary acts are preceded by a slow negative shift in electrical potential recorded at the scalp known as the readiness potential (RP) and that this shift can precede the act by up to one second or more. More significantly, the neurophysiologist Benjamin Libet found that, for simple spontaneous acts such as flexing the wrist or fingers, RP even preceded the experienced wish to flex the wrist or fingers by around 350 milliseconds, suggesting that the brain is preparing to act even before the conscious wish to act appears!

Given its radical consequences, this finding and the experimental methods used to obtain it have been the subject of considerable discussion and debate. The broad consensus, however, is that the findings are not just artifacts. They were confirmed, for example, by Patrick Haggard and Martin Eimer who went on to investigate the preparedness of the brain to act with either the left or right hand, indexed by the lateralized readiness potential (LRP), a negative shift in electrical potential in the cerebral hemisphere on the opposite side to the active hand, when compared to the electrical potential in the one on the same side. Again, as with RP, LRP occurred before the conscious wish to move a given hand, although in this case by around 100 milliseconds (ms).

In an attempt to find a role for the conscious experiences of willing themselves, Libet pointed out that although conscious wishes follow the RP by around 350 ms, they precede the act by around 150 ms, time enough to veto the wish. So he suggested that the ability to veto the wish is the function of conscious volition (rather like a conscious Freudian ego controlling the unconscious id). However, a decision not to act (after a readiness to do so) can be shown to have its own antecedents. Using go or no-go tasks where subjects were given a signal to either press a button or withhold pressing that button, Gethin Hughes found that response inhibition could be preconsciously influenced in various ways. For example, he found that a signal not to press the button presented 100 ms before the conscious no-go signal, produced earlier response inhibition, even when the preconscious signal was prevented from entering consciousness by a masking signal that obscured it. To solve the problem that there is no overt behavior in the no-go condition, the onset of response inhibition was determined from the onset of the no-go N2, a negative going cortical potential measured over frontally placed electrodes that acts as an index of response inhibition, occurring about 200 ms after a conscious cue not to act appears.

Where Free Will Fits Into Mental Processing

Given such findings, is conscious free will an illusion? In assessing this, it is important to note that free will is not a special case. As Max Velmans has pointed out, just as conscious wishes and decisions follow the neural processing required to generate them, conscious percepts follow the perceptual processing required to produce them, conscious thoughts follow the cognitive processing required to produce them and so on. In these cases, the conscious phenomenology that results from brain processing represents the outcome of that processing without itself being that processing—and it is indeed illusory to confuse the products of mental processing with the processing itself. Such conscious experiences can nevertheless give useful information about their mental antecedents. Visual experiences give useful information about the visual features of the world that have been processed; thoughts in the form of inner speech provide useful information about the nature of prior cognitive processing and so on. Similarly, feelings of being able to choose among alternatives can accurately reflect the operation of preconscious decision-making processes constrained by inner goals and needs, social and physical constraints, and so on—supporting a form of what Velmans refers to as “preconscious free will.”

Max Velmans

See also Consciousness and the Unconscious; Freedom of Action; Voluntary Action, Illusion of

Further Readings


Successful prediction is an essential activity in scientific and lay life, and this entry traces the nature and lessons of successful prediction. Suppose you have to make a prediction about some target property (e.g., a child’s adult height or whether a prisoner will recidivate) on the basis of \( n \) lines of evidence (e.g., the child's height at 3 years of age, the heights of the child's parents). It doesn't matter how those lines of evidence were acquired—whether via a clinician's judgment or a mechanical procedure (e.g., a computer-graded aptitude test). To arrive at a prediction, one must weigh and combine those lines of evidence and come to a judgment. Clinical prediction is any prediction in which the weighing and combining of evidence is done by an expert human. Actuarial prediction is a purely mechanical procedure in which the weighing and combining of evidence is done algorithmically, via a transparent formula that could be applied equally well by a trained clinician or by a clerical worker.

The literature comparing clinical and actuarial prediction begins with Paul Meehl's 1954 classic, *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Meehl reports on 20 studies in which actuarial predictions, based on very simple actuarial rules, were more accurate than clinical predictions. Since the publication of what Meehl called his “disturbing little book,” psychologists have developed many successful actuarial models or statistical prediction rules (SPRs).

The literature comparing clinical and actuarial prediction begins with Paul Meehl's 1954 classic, *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Meehl reports on 20 studies in which actuarial predictions, based on very simple actuarial rules, were more accurate than clinical predictions. Since the publication of what Meehl called his “disturbing little book,” psychologists have developed many successful actuarial models or statistical prediction rules (SPRs).

**The Golden Rule of Predictive Modeling**

There is now overwhelming evidence for “the golden rule of predictive modeling”: When given identical evidence, well-constructed SPRs predict at least as reliably, and typically more reliably, than human experts. The most decisive case for the golden rule has been made by William Grove and Paul Meehl, who reported on an exhaustive search for studies comparing SPR predictions to human predictions in which (a) the humans and SPRs made predictions about the same specific cases and (b) the SPRs never had more information than the humans (although the humans often had more information than the SPRs). Their research yielded 136 studies comprising 617 distinct comparisons between SPR and human predictions. These studies covered a wide range of predictive efforts, including medical and mental health diagnoses; treatment prognoses, recommendations, and outcomes; descriptions of personality; adjustment to institutional life in both the military and correctional facilities; success in training or employment; socially relevant aggregate behaviors such as business performance; and many others. Of the 136 studies Grove and Meehl analyzed, 64 clearly favored the SPR, 64 showed approximately equivalent accuracy, and eight clearly favored the human predictor. The eight studies in which the human clinician outperformed the SPR appeared to have no common characteristics and therefore are not indicative of a unique domain in which expert prediction reliably beats statistical prediction; they are simply outliers (given 136 chances, the better reasoning strategy is bound to lose sometimes, after all).

**What Explains the Golden Rule?**

There are many sophisticated prediction models on the market (e.g., neural networks, naïve Bayes classifiers, classification and regression trees, support vector machines). The most sophisticated of these are optimizing models: They begin with a data set that consists of various lines of evidence that are correlated with the target property and then they employ sophisticated mathematical techniques with the aim of weighing those different lines of evidence so as to best predict new data. Perhaps the most common optimizing model is the proper linear (or regression) model. An intuitive way to understand proper linear models is to suppose we are trying to predict a target property (a person’s weight) on the basis of a single line of evidence (the person’s height). Suppose we take a set of data (for example, the heights and weights of a large number of people) and plot them on a graph. When doing so, a proper linear model will draw a straight line on the graph that comes closest to all the data points. This line can be described by the formula, \( y = k + cx \), where \( y \) stands for weight, \( x \) stands for height, \( k \) is the \( y \)-intercept,
and $c$ is the slope of the line. In most real-life cases, however, we will have more than just one predictor cue, and so we need a multiple regression equation. The trick to building a proper (optimizing) linear model given complex data involves choosing the coefficients ($c_n$—in the above example, the weights) so that the model best fits the data in the training set. Such equations have the form

$$y = k + c_1x_1 + c_2x_2 + c_3x_3$$

What explanation can be given for the reliability of the golden rule? The obvious answer seems to be that unaided experts cannot possibly hope to be as accurate as optimizing models. On this explanation, unaided experts simply cannot construct and implement an optimizing model in their heads because they can’t absorb and process all the available evidence, and even if they could, they can’t assign optimal weights to the different lines of evidence, and even if they could do that, they cannot solve the model’s complex formula to arrive at the model’s prediction. However, this explanation can’t be right because, in practice, some fairly simple nonoptimizing models are also more accurate than human experts. Consider three such models:

1. **Bootstrapping models:** The bootstrapping model is a proper linear model of a person’s judgments about a target property but an improper linear model (a model that does not best fit the available data) of the target property itself. The bootstrapping model is built to, in essence, predict the human expert’s prediction. And it will, from time to time, be wrong about what the expert will predict. But when it is wrong about the expert, it’s more likely to be right about the target property!

2. **Unit weight models:** The unit weight model assigns equal weights to standardized predictor cues so that each input has an equal bearing on the final prediction. Given the success of unit weight models, an unweighted sum of a few of the most predictive variables will tend to be preferable to more complicated regressions equations. What is so surprising is how simply these formulas can be calculated: All you need is knowledge of what those most relevant variables are and then be able to add them all up.

3. **Randomized models:** The bootstrapping models discussed above can also be altered to produce a random linear model. In random linear models, there is no attempt to assign optimum weights to variables. Instead, variables are given random weights—with one important qualification: All the cues are defined so they are positively correlated with the target property. Even given this stipulation, random linear models are still as reliable as the proper models and more reliable than human experts.

A fascinating principle underlies these seemingly counterintuitive findings: The flat maximum principle says that for a certain class of prediction problems, as long as the signs of the coefficients are right, any one linear model will predict about as well as any other. It is important to recognize that the flat maximum principle is restricted to certain kinds of problems in which the following conditions obtain:

1. The judgment problem must be difficult and one in which no proper model will be especially reliable, because the world is messy.
2. The evidential cues in the problem must be reasonably predictive. For example, the best cues for predicting academic performance (GPA, test scores) are reasonably predictive. Certainly, a reasonably predictive cue is one that is at least more reliable than chance.
3. The evidential cues must be somewhat redundant. For example, people with higher GPAs tend to have higher test scores.

Given how common these circumstances are, it is in fact not uncommon for the improper unit weight models to be more reliable than the proper models.

**The Golden Rule in Practice**

The best way to get a sense of the power of actuarial prediction is to consider some examples.

1. Given a patient’s marital status, length of psychotic distress, and the patient’s insight into his or her condition, an SPR was more reliable at predicting the success of electroshock therapy than medical and psychological staff.
2. Criminologists were less reliable in predicting criminal recidivism than an SPR based on criminal and prison records.

3. An SPR was more reliable than clinical psychologists in diagnosing patients as either neurotic or psychotic, initially using the basis of a Minnesota Multiphasic Personality Inventory (MMPI) profile and then assessing diagnostic accuracy on follow-up. Even when psychologists were given the SPR’s results before they made their predictions, they were still less accurate than the SPR.

4. SPRs predict academic performance (in terms of graduation rates and GPA at graduation) better than admissions officers at selective colleges, law schools, medical schools, and graduate school in psychology. This discrepancy holds even when the admissions officers have access to considerably more evidence than the models.

5. SPRs predict loan and credit risk better than bankers. SPRs, not human experts, are now standardly used by banks to make loans and by credit card companies to approve and set credit limits for new customers.

6. SPRs predict newborns at risk for sudden infant death syndrome (SIDS) with much greater accuracy than human experts.

7. An SPR predicts the quality of the vintage of a red Bordeaux wine better than expert wine tasters who are able to swirl, smell, and taste the young wine.

8. An SPR correctly diagnosed 83% of cases of progressive brain dysfunction on the basis of input cues in the form of intellectual tests, while clinicians working from the same data did no better than 63%. When given the results of the actuarial formula, clinicians still did worse than the model, scoring no better than 75%.

9. An SPR outperformed experienced clinicians as well as a nationally renowned neuropsychologist in predicting the presence, location, and cause of brain damage.

10. When predicting violence in a legal setting, one will actually be more reliable than forensic psychologists simply by predicting that people will not be violent. In addition, SPRs are more reliable than forensic psychologists in predicting relative likelihoods of violence.

As these examples suggest, actuarial prediction remains the province of specialized domains such as criminal recidivism or credit risk executed by professionals trained in predictive modeling. If actuarial methods are to actualize their potential to improve clinical and lay judgment, they will have to be made more widely available, easier to use, and applicable to a broader range of problems.

Michael A Bishop and J. D. Trout

See also Decision Improvement Technologies; Modeling Causal Learning; Scientific Reasoning; Thinking

Further Readings


**Production of Language**

Producing a linguistic expression involves retrieving a set of words, arranged in a grammatical hierarchically organized sequence so as to convey speakers’ intended thoughts. This is accomplished through a series of processing steps. A speaker begins an act
of language production by deciding what to say: message formulation. For example, a cat owner might want to tell you his or her pet happened to eat an unspecified arachnid. Then a speaker must decide which words to use to express that message, lexical selection, and then retrieve those words, lexical retrieval. The cat owner might choose the words my cat, ate, and a spider. A speaker must also assign these words to roles that convey who did what to whom, function assignment. The cat owner might assign “my cat” to the subject function and “a spider” to the object function. At least in spoken languages such as English, words can only be produced one after another, so a speaker must then use the principles of the grammar of the speaker's language to order the role-assigned words, constituent assembly. In English, the subject goes before the verb and the object after, so the cat owner's sentence will use the sequence “my cat ate a spider.” With words and their order (at least partly) determined, a speaker can send a plan off to phonological encoding so that the sound of an utterance can be formulated, followed by articulation so that a signal is actually generated for an audience. Additionally, most approaches to language production allow monitoring—that is, assessing formulated speech (before or after articulation) for adequacy and accuracy.

To be precise, though it is convenient to present these processing components as operating in a strictly sequenced or staged fashion, it is not clear that the language system works this way. Certainly, it is not the case that each component must finish its tasks completely for an entire sentence before the next stage can start. Rather, production is incremental in the sense that once some initial part of an utterance has been formulated at one level of processing, that part can be sent for processing at the next level of processing as upcoming parts of the utterance are formulated at the first level. For example, once a speaker has retrieved the words my and cat, they can be sent off for phonological encoding as the word ate is selected. Such incrementality allows speakers to start their utterances sooner and, more generally, permits some parallel processing so as to make production more efficient. A more controversial aspect of production planning, whether processing is strictly staged or is more free-flowing, is discussed below where lexical selection and retrieval are detailed further.

Each of these processing components is discussed in turn.

**Message Formulation**

The first step a speaker must take to produce a linguistic expression is to formulate a message. Based on logical analysis, message formulation is seen as proceeding through macroplanning and microplanning. Macroplanning involves determining a goal for an utterance and choosing the information needed to express it. Microplanning involves taking a particular perspective on the meaning to be expressed and determining the more versus less important elements of that meaning. The point of all these steps is to formulate a complete, dynamic representation of the information a speaker aims to convey in words and phrases.

One aspect of message formulation that has been heavily studied concerns how we adopt the perspective of our addressees to choose the right bits of meaning to linguistically encode. Specifically, any informative utterance includes some information a speaker's addressee does not yet know—what the speaker wants to convey—described in terms of information the addressee already knows. An utterance out of the blue such as “Steve ran the Boilermaker” will likely be uninformative, but an utterance such as “My friend Steve ran a 15 kilometer race in upstate New York called the Boilermaker” will likely be more informative. The reason is the latter utterance describes new information (Steve, the Boilermaker) by using information the addressee already knows. An utterance out of the blue such as “Steve ran the Boilermaker” will likely be uninformative, but an utterance such as “My friend Steve ran a 15 kilometer race in upstate New York called the Boilermaker” will likely be more informative. The reason is the latter utterance describes new information (Steve, the Boilermaker) by using information the addressee already knows. An utterance out of the blue such as “Steve ran the Boilermaker” will likely be uninformative, but an utterance such as “My friend Steve ran a 15 kilometer race in upstate New York called the Boilermaker” will likely be more informative. The reason is the latter utterance describes new information (Steve, the Boilermaker) by using information the addressee already knows. An utterance out of the blue such as “Steve ran the Boilermaker” will likely be uninformative, but an utterance such as “My friend Steve ran a 15 kilometer race in upstate New York called the Boilermaker” will likely be more informative. The reason is the latter utterance describes new information (Steve, the Boilermaker) by using information the addressee already knows. An utterance out of the blue such as “Steve ran the Boilermaker” will likely be uninformative, but an utterance such as “My friend Steve ran a 15 kilometer race in upstate New York called the Boilermaker” will likely be more informative. The reason is the latter utterance describes new information (Steve, the Boilermaker) by using information the addressee already knows. An utterance out of the blue such as “Steve ran the Boilermaker” will likely be uninformative, but an utterance such as “My friend Steve ran a 15 kilometer race in upstate New York called the Boilermaker” will likely be more informative. The reason is the latter utterance describes new information (Steve, the Boilermaker) by using information the addressee already knows. An utterance out of the blue such as “Steve ran the Boilermaker” will likely be uninformative, but an utterance such as “My friend Steve ran a 15 kilometer race in upstate New York called the Boilermaker” will likely be more informative. The reason is the latter utterance describes new information (Steve, the Boilermaker) by using information the addressee already knows.
ground. The other type describes the processing strategies speakers might use to compute common ground. Of the first type, speakers might assume that if a fact had been mentioned in the presence of some interlocutors then that fact should be ascribed to common ground. Similarly, if some noticeable feature is in the immediate environment of interlocutors, it could be ascribed to common ground. Of the second type, speakers might track the statistical reliability of different sorts of cues to common ground (e.g., whether a partner’s visual perspective tends to indicate that they do or do not know about some element of the environment), or they might assume that their addressees know everything they themselves know except for explicitly excluded privileged information, or they might specifically track relatively simple bits of information that can be used to ascribe knowledge to specific addressees. In all these cases, the point is to be able to know what knowledge speakers and addressees share so that it can be relied on to successfully convey new knowledge.

**Lexical Selection and Retrieval**

Given a message, speakers must retrieve a set of content words—words describing entities, states, and actions—that can convey the meaning specified in the message. This is typically thought to proceed through two steps: First, speakers must find and select the words that convey the intended meaning; then speakers must retrieve the phonological features that represent the sounds of the words.

Given a particular meaning to convey, multiple lexical forms similar in meaning to the to-be-expressed meaning are accessed. Each of these becomes accessible in proportion to its degree of meaning similarity to the to-be-expressed meaning. This can be thought of as a kind of search: Given that a given to-be-expressed meaning will not necessarily correspond neatly to an individual word, a process of accessing a range of candidates to varying degrees allows an appropriate word to be found. This also allows other factors to influence the selection process, possibly including the accessibility or suitability of the phonological properties of the potential word.

According to most (though not all) approaches to word production, a competitive process then operates to select the to-be-produced word. This will make it so that if many possible words are relatively more accessible, the selection of a target will happen more slowly, but if only one word is highly accessible, selection will be quicker. For example, a speaker might want to name an unusual bug she or he saw. Visualizing the bug will cause its meaning features to become represented in the speaker’s message, which in turn will lead the words that can express those meaning features to be accessed. If the bug looked much like a beetle but also somewhat like a tick or a roach, then the lexical representation of *beetle* will be accessed to a greater degree and the lexical representations of *tick* and *roach* to lesser degrees. The more (the speaker represents that) the bug looked like the beetle and the tick and the roach, the closer these accessibility levels will be and (if selection is competitive) the longer it will take for selection to occur. If the bug looked more like a beetle and less like a tick or roach, then the accessibility level of *beetle* will be much greater than the accessibility levels of *roach* and *tick*, and so selection will happen more quickly.

It is important to note that this selection process (competitive or otherwise) is restricted. For example, when speakers make speech errors where they select a wrong word, an intended noun is very often replaced with another noun, a verb with another verb, and so forth. This suggests that lexical selection operates within syntactic category; for example, only nouns are considered if a noun is to be selected. It may be that other sorts of categories similarly restrict lexical selection (e.g., whether a word is at an intended level of abstraction; for bilingual speakers, what language a word is in).

Production processing may be discrete, if selection must complete before retrieval begins, or cascaded, if retrieval can begin for accessible forms even before they are selected. Also, if lexical selection and retrieval is cascaded, it may allow feedback, if the retrieval of phonological properties can affect lexical selection. There are good reasons to allow cascading (it presumably speeds retrieval if it is begun sooner) or to forbid it (there is no point retrieving the phonological features of forms that will not be selected), and there are good reasons to allow feedback (it may be good to allow sound properties to affect selection) or to forbid it (it may be bad to allow sound properties to affect selection). A range of evidence suggests that lexical selection and retrieval is indeed cascaded; whether lexical selection and retrieval allows feedback is less certain.
According to some views, lexical retrieval begins with the retrieval of a whole word (or more precisely, whole morpheme—the atomic unit of meaning in language) representation. Then individual speech sounds—phonemes or segments—are retrieved. Phonemes are arranged into syllables (a unit of sound including a vowel and some of its immediately surrounding consonants), which in turn are specified for metrical properties such as whether they should be pronounced with more stress (louder, longer) or less (e.g., different metrical patterns need to be retrieved to produce the noun or verb forms of a word such as record: REcord vs. reCORD). At this point, language production processes have completed their lexical job, and speech and motor processing begins.

Function Assignment and Constituent Assembly

Utterances do not just convey the individual bits of content that words express. Utterances also convey the relationships among those bits of content—the “who did what to whom” of language. To do this, to-be-produced content words must be assigned to grammatical roles that indicate what relationship they bear to some event. This is the job of function assignment: to assign content words to grammatical functions such as subject and object so that their role with respect to an event can be expressed.

Words and their grammatical functions must somehow be outwardly encoded in the to-be-produced utterance. Spoken languages use different strategies for expressing grammatical functions such as subject and object. Languages such as English largely use the relative positions of words in utterances to convey grammatical functions; these languages are thus termed word order (or more precisely, fixed word order) languages. In contrast, languages such as Japanese largely use special morphemes (usually suffixes) to convey grammatical functions; these languages are sometimes called case-marking languages. So “Steve-ga Boilermaker-o hasitta,” where the -ga and -o suffixes convey subject and object, respectively.

The complex task of determining the relative order of words is carried out by constituent assembly processes. This is made complex by the fact that multiple relationships can be embedded in a to-be-expressed thought, and all these must be encoded in a simple linear sequence of words (possibly with case markers, in a case-marking language). For example, “My friend Steve ran the Boilermaker” conveys a relationship between running, Steve, and the Boilermaker but also between friend and Steve, and all these must be appropriately nested with respect to one another. To determine the right way to convey such nestings, constituent assembly processes in some cases must consult representations of the principles of the speaker’s grammar to determine what the linear sequence of words should be so that (in English) the adjective ends up preceding the noun, the subject preceding the verb, and so forth. But the grammar does not determine all ordering of words. As noted, case-marking languages such as Japanese do not use word order to indicate subject and object roles, and so constituent assembly is free to order subjects and objects without regard to grammatical principles (and because of this, case marking languages are sometimes called free word order languages). So “Steve-ga Boilermaker-o hasitta,” and “Boilermaker-o Steve-ga hasitta” both convey the same (overall) meaning. And even in English, the order of nouns in a conjunction is not grammatically determined (“Steve and Kim” or “Kim and Steve” are both grammatical). In these cases, other factors will influence relative ordering, such as the prominence or personal importance of the to-be-ordered items or how easily the sentence material can be retrieved from memory.

Different factors affect function assignment and constituent assembly, revealing the nature of these processes. Speakers tend to repeat the hierarchical structures of sentences they recently heard or said, a phenomenon termed structural or syntactic priming. Structural priming has been profitably used to determine the representational nature of the hierarchical frames that are enlisted for production. Speakers also will tend to produce sentence structures that allow easily accessed material to be produced sooner and harder to access material later. These are often termed accessibility effects and may be motivated by efficiency (saying words as they are retrieved reduces memory demands). Agreement, whereby different parts of a sentence must systematically covary (e.g., the subject and verb in English agree in number),
Production of Language has been heavily investigated. Research on agreement has shown that sentence production processes enlist formal, abstract representations that underlie the relevant agreement properties (e.g., singular vs. plural in English).

Phonological Encoding
Above, the processes involved in retrieving the phonological properties of individual words were described. But words are of course produced as parts of longer utterances, and those utterances have their own phonological properties. Most prominently, extended utterances have *prosody*—the melody and rhythm of an utterance that conveys different sorts of information. Specifically, prosody can convey syntactic information, including where phrase boundaries are between words in a sentence; whether a sentence is a declarative, an interrogative, or something else; emotional information, such as whether a speaker is excited or sullen; and so forth. Consequently, the mechanisms responsible for formulating prosody must take into account all these sorts of information.

An important aspect of prosody that must be computed concerns how long the words and pauses should be in an utterance. Production research suggests that prosodic production mechanisms specify at least some of the duration properties of the “slots” words are to be spoken in, and this is done relatively independently of the content of the words themselves. This implies that when a word is shorter, speakers will compensate by producing a longer pause after it and vice versa. This is an elegant demonstration of the relative independence of different production mechanisms, here, between prosodic formulation and lexical retrieval mechanisms.

Monitoring
The task of language production is not completed when an utterance is fully formulated and ready for articulation. Production mechanisms also engage in monitoring, whereby the adequacy and accuracy of a formulated expression is checked. Formulated utterances are monitored both before they begin to be articulated, based on some form of inner speech, and after they are articulated, through standard language comprehension mechanisms. Evidence for the former is that speakers will sometimes halt the articulation of an erroneous utterance quickly—so quickly that it is not possible that the speaker heard the error (externally) and halted production on that basis. Monitoring based on external speech is evident whenever speakers notice that they have produced some utterance they did not intend or when they realize an utterance did not “come across right.”

Different mechanisms might be responsible for carrying out monitoring. One common proposal, the *perceptual loop* hypothesis, is that formulated speech (both inner and external) is monitored with the same comprehension mechanisms we use to comprehend others’ (external) speech. The idea is that formulated speech is comprehended to the level of meaning and then the comprehended and to-be-expressed meanings compared; if they differ in any important way, the speaker can stop and reformulate. But other monitoring mechanisms (and more generally, error-detecting mechanisms) may operate in addition to or instead of the perceptual loop. For example, one strategy for monitoring for errors is to detect when some representation at a later level of processing has been selected even though its antecedent representation has not. If a speaker intends to say “cat” but formulates “lat” instead, production mechanisms can detect the error if they are sensitive to the fact that the *J* segment was accessed for production even though no word including an *J* was selected during lexical selection and retrieval.

Conclusions
The ease with which we produce speech belies the complexity of the cognitive mechanisms underlying this ability. This allows us to convey our thoughts to other members of our species with a level of detail and breadth that would not otherwise be possible. This ability, one of only a handful unique to humans, is critical to our survival and success as a species.

*Victor S. Ferreira*

See also Conversation and Dialogue; Language Production, Agreement in; Language Production, Incremental Processing in; Natural Language Generation; Perspective Taking in Language Processing; Planning in Language Production; Prosody in Production

Further Readings

**PROSODY IN PRODUCTION**

Although the term *prosody* has been used to describe a wide variety of linguistic and extralinguistic phenomenon, it can, roughly, be defined as the acoustic aspects of an utterance that vary independently of the phonology of the utterance’s words. Prosody plays a clear role in conveying information about discourse, syntax, and pragmatics to listeners. However, prosody is also linked to processes related to language production. This entry discusses what prosody is and how it is related to speech production.

### Prosodic Structure

Prosody includes acoustic information associated with rhythm, intonation, pausing, and accents (emphasis). Prosodic information is typically conveyed through changes in fundamental frequency ($f_0$), which correlates with the perception of pitch, and changes in intensity, which correlates with the perception of loudness. It is also linked to pausing and the lengthening and reduction of individual words.

Prosody can vary between otherwise identical utterances in ways that influence the interpretation of a sentence. For example, consider (1), where capitalization indicates an accent and “//” indicates a break in the speech stream.

(1a) Ketchup is a vegetable.
(1b) Ketchup is a vegetable?
(1c) KETCHUP is a vegetable.
(1d) Ketchup is a VEGETABLE.
(1e) The cop saw // the spy with the telescope.
(1f) The cop saw the spy // with the telescope.

The sentence in (1a) will sound like an assertion if it is produced with a fall in $f_0$ across the sentence, but it will sound like a question if it is produced with a rise (1b). These two productions convey different speaker attitudes about the truth of the utterance. In (1c), accenting *ketchup* conveys that it and not, say, mustard is a vegetable. Similarly, (1d) implies that ketchup is a vegetable and not a fruit. Breaks in the speech stream, which are also called prosodic boundaries, can be signaled by a pause, a change in $f_0$, or lengthening of the preboundary word. A prosodic boundary after the verb *saw* in (1e) biases the listener toward an interpretation in which the spy has a telescope whereas a break after *spy* biases the listener toward an interpretation in which the cop has the telescope. These examples illustrate that prosody can influence the interpretation of an utterance even if the words in the utterance do not vary.

Prosodic features such as prosodic boundaries, accents, rhythm, and pitch contours are represented at an abstract phonological level of representation called prosodic structure. There is a great deal of controversy over the nature of this representation. There is disagreement over whether there are...
different types of prosodic phrases and how they might differ, whether and how these phrases are represented hierarchically, whether accents differ in form and meaning, how intonational contours are structured, and the nature of the interface between prosodic structure and other aspects of language.

**Production**

Traditionally, it has been assumed that the role of prosody is to convey pragmatic, discourse, and syntactic information to listeners. However, a growing body of evidence suggests that there might also be a link between speaker internal production processes and prosody: Speakers are more likely to produce prosodic boundaries and accents at points of production difficulty.

For example, the likelihood of producing a prosodic boundary either before or after a syntactic constituent such as the subject of a sentence is directly correlated with the constituent’s length.

(2a) The judge who the reporter for the newspaper ignored // fired the secretary.
(2b) The judge who the reporter ignored (//) fired the secretary.

In (2), a boundary is more likely to follow the subject of the sentence in (a) than (b) because the subject in (a) is longer. Boundaries may provide speakers with time to plan upcoming structure or recover after encountering production difficulty.

A question of much debate in this literature is whether speakers purposely produce boundaries in locations that are helpful to listeners, particularly ambiguous sentences such as the ones in (1e) and (1f), or whether they produce boundaries to help themselves. While some researchers have found that speakers do not use boundaries to consistently disambiguate syntactically ambiguous sentences for listeners, others have found that they do. Many of the experiments investigating this question have used tasks in which a speaker must instruct a listener to manipulate a set of objects.

(3) Tap (a) the frog (b) with the flower.
(4) Put the dog (a) in the basket (b) on the star.

Jesse Snedeker and John Trueswell have found that in sentences such as (3), speakers disambiguate the sentence using boundaries at positions (a) or (b) only if they are aware of the ambiguity. However, Tanya Kraljic and Susan Brennan have found that in sentences like (4), speakers disambiguate the sentence whether they are aware of the ambiguity or not. One critical difference between (3) and (4) is utterance length: Sentence (4) is longer. If prosodic boundary placement is partly influenced by production constraints, then one would expect boundaries to occur more often in long, difficult sentences than short, easy sentences, independent of the needs of the listener, and this appears to be the case. However, the extent to which boundary placement is speaker or listener centered is currently under debate.

This debate also extends to the literature on accents. Word lengthening, a correlate of accenting, co-occurs with disfluencies and boundaries, suggesting that production difficulty may underlie both. Words that are unpredictable, informative, and low frequency are more likely to be accented than those that are not. These words are likely to be difficult to produce although it is unclear whether prominence in these cases is the result of production difficulty or whether speakers mark words that might be difficult for the listener in order to facilitate comprehension. This question is one that researchers are currently wrestling with.

Overall, it is clear that prosody can signal information about syntax, discourse, and pragmatics, and recent work suggests that prosody may also play an important role in language production.

_Duane G. Watson_

See also Planning in Language Production; Production of Language; Speech Perception

**Further Readings**


After a neuron fires, it becomes temporarily unable to fire again, no matter how great the excitation. In 1931, Charles Telford wondered whether higher level cognitive processes might be subject to an analogous limitation. That is, after one act of cognition, cognitive-processing mechanisms might become much less excitable during a prolonged recovery phase, leading to slow responses. To test this conjecture, Telford asked his subjects (psychology graduate students) to press a telegraph key as fast as possible whenever they heard a sound. The time between consecutive sounds was either 0.5, 1, 2, or 4 seconds, determined randomly on each trial. Although the task was remarkably simple, response times increased from 0.241 sec to 0.335 sec as the time between tones decreased from 1 sec to 0.5 sec (see Figure 1). These findings appeared to directly support Telford’s refractory period hypothesis.

The phenomenon Telford discovered—which became known as the psychological refractory period effect—has been extremely influential. It is of great practical interest, because it bears on the multitasking difficulties faced by human operators of complex systems (which today would appear to include virtually the entire human adult population). His discovery is also of great theoretical interest, because it points to a fundamental limitation in the human ability to process information. An enormous amount of research has been devoted to uncovering the nature of that fundamental limitation.

In addition, his experimental paradigm has been adopted in many hundreds of subsequent experiments and is, in fact, still quite popular today (albeit using far more sophisticated equipment for presenting stimuli and measuring responses).

Although Telford’s paradigm and empirical findings were extremely influential, his specific theory has long since been abandoned. A refractory period suggests that cognitive mechanisms are temporarily sluggish—or less responsive—immediately following an act of cognition. Research has revealed no evidence for sluggishness. Although the analogy with the neuronal refractory period seems misguided, the misleading label psychological refractory period (PRP for short) became securely attached to this effect. This entry describes modern theoretical accounts of the PRP effect, efforts to test between them, and implications for dual tasking in the real world.

Theories of the Psychological Refractory Period Effect

Instead of a refractory period, contemporary theories propose that cognitive mechanisms are continuously and fully engaged yet have difficulty serving more than one task at a time. The two most prominent accounts of the PRP effect are capacity-sharing models and bottleneck models.

Capacity-Sharing Models

According to capacity-sharing models, humans have a limited pool of capacity to divide between concurrent tasks. An analogy would be a general that divides one large force into two smaller forces, each simultaneously pursuing a different objective. On this view, tasks can operate in parallel. But because they receive only a fraction of the available capacity, they will operate more slowly than is possible under single-task conditions.

Which mental operations require a share of the limited pool of resources? The evidence suggests that multiple perceptual processes can operate in parallel, at least under favorable circumstances (imagine a loud tone presented at the same time as a letter on the computer screen). It also seems clear that people can generally execute simple responses in parallel, as in walking and talking. These considerations suggest that the processes requiring capacity lie somewhere in between perception and response execution, in
what is referred to as central processing. A clear-cut example of a central process is response selection—deciding what response is warranted given the stimuli in the environment.

**Bottleneck Models**

Bottleneck models take the rather extreme view that certain mental processes simply cannot operate on more than one task at a time. As an analogy, a bank teller can usually serve only one customer at a time. If two customers arrive at the same time, one of them will need to wait for the other to finish.

Alan Welford, in 1952, was the first to specifically attribute the PRP effect to a bottleneck in central mental operations (see Figure 2). In other words, while any central operation for the task arriving first (Task 1) is underway, all central operations for the second task (Task 2) must necessarily wait. Again, central operations involve deciding how to respond (or not respond) to a given stimulus. Welford showed that this model could account for a wide range of data from the PRP paradigm.

A few decades later, Harold Pashler revived interest in the central bottleneck model by demonstrating that it makes several specific and counterintuitive predictions. For example, the effects of degrading the Task 2 stimulus should actually be much smaller under dual-task conditions than single-task conditions. Consider the concrete example of a Task 2 that involves classifying a letter presented on a computer screen. Dimming this Task 2 letter might prolong perceptual processes by about 50 milliseconds (ms), which should in turn increase single-task response times by 50 ms. But surprisingly, this same 50-ms perceptual delay should have little or no effect on response time in a PRP experiment with a short delay between the Task 1 and Task 2 stimulus onsets. The reason is that Task 2 perceptual operations have ample time to finish while Task 2 central operations wait for Task 1 central operations to finish. As an analogy, a delay in completing a deposit slip will delay your departure from the bank when there are no other customers (single task). But if you must wait for the teller to first help another customer (dual task for the bank teller), any delay in completing the deposit slip will likely be inconsequential. You can simply complete the deposit slip while waiting in line.

This prediction is counterintuitive because one would normally expect that the negative consequence of increasing task difficulty would be magnified under dual-task conditions. Nevertheless, many experiments have confirmed this prediction of the central bottleneck model. The effects of dimming a Task 2 stimulus or superimposing a noise mask are virtually absent when the delay between the Task 1 and Task 2 stimulus onsets is short. The successful predictions of the central bottleneck model support the hypothesis that central stages often do operate serially (one at a time) in the PRP paradigm.

Data from many PRP experiments are consistent with serial central processing—that is, allocation of attention entirely to Task 1 central operations then entirely to Task 2 central operations. It has been argued, however, that simultaneous sharing of capacity is possible when the conditions favor it (e.g., when tasks are assigned equal priority). Although this debate has not yet been resolved, both sides of the debate appear to agree that PRP effects stem mainly from a limited capacity to carry out central mental operations.

Are There Exceptions to the Psychology Refractory Period Effect?

PRP effects are surprisingly robust. They have been reported with a wide variety of different judgments, even very simple ones. Importantly, they have been reported even when the two tasks are maximally different, with distinct inputs (seeing vs. hearing) and distinct outputs (speaking vs. moving the hands). These findings hint at a pervasive problem that could only worsen with more complex, real-world tasks. Driving a car, for example, is vastly more complicated than Telford’s task of pushing a key in response to a sound.

But is there no way to eliminate the PRP effect? Many investigators have attempted to eliminate the PRP effect by using extremely natural tasks, such as repeating a spoken word or moving a joystick in

---

**Figure 2** The central bottleneck model
the direction of an arrow. In such cases, the stimulus might strongly activate the corresponding response without any assistance from central mental resources. Perhaps not surprisingly, PRP effects are reduced with such tasks. However, stubborn residual PRP effects have often been reported, hinting that even these tasks might be subject to resource limitations.

Other investigators have taken a different approach. Instead of using inherently easy tasks, they have used somewhat more difficult tasks but allowed subjects the opportunity to master them. Obviously, tasks are performed faster with practice, so the PRP effect should inevitably decrease. The critical question is whether practice can eliminate the PRP effect altogether and, more to the point, whether practice eliminates competition for mental resources. This type of automaticity with practice has in fact been clearly documented in a few recent studies. Most of these demonstrations, however, have thus far involved relatively trivial tasks, involving a very small number of stimuli and responses.

What about highly practiced real-world tasks, for which the array of possible stimuli is essentially infinite? Consider, for example, driving while talking on a cell phone. Despite being extremely well practiced, these tasks nevertheless interfere with each other. Such interference has been documented both in analyses of actual accident reports and in numerous experiments with simulated driving tasks. It has been estimated that cell phone use quadruples the frequency of accidents and is as detrimental to driving as a blood-alcohol level of .08 (considered legally intoxicated in some states). Interestingly, this interference is not ameliorated by using hands-free phones rather than handheld phones. This finding hints that the problem lies in interference between central processes (i.e., cognitive distraction) rather than in competition between tasks for the hands. Ignoring these consistent research findings, lawmakers typically outlaw the use of handheld cell phones while driving yet permit the use of hands-free cell phones.

Another interesting exception to the PRP effect occurs when the second task merely requires moving the eyes (whose position is monitored using special equipment) to a specified type of stimulus, such as any red object. Eye movements might evade the PRP effect because they are quasi-reflexive, supported by special neural circuitry. An alternative view is that fixating one’s eyes on a target is a highly practiced action, and it is the unusually high levels of practice that enabled subjects to escape the PRP effect.

Central Bottlenecks: Structural Limitation or a Strategic Choice?
The central bottleneck model implies that people cannot perform central operations on more than one task at a time because of some cognitive resource limitation. This limitation is often described as a single-channel cognitive mechanism that can process only one task at a time. Strictly speaking, however, most data merely show that people do not perform central operations at the same time. Perhaps people are entirely capable of parallel central processing but typically choose serial processing as a matter of strategy.

This issue remains controversial. At low levels of practice, attempts to induce subjects to choose a parallel processing strategy, and eliminate the PRP effect, have not been very successful. Parallel processing does seem to be possible at higher practice levels, as noted above, although it is not clear that the transition reflects a voluntary choice, per se. Ultimately, the strategic versus structural debate might not have a simple answer. For instance, a structural limitation in central mechanisms might lead to a strategic choice to perform central operations serially (one at a time). Computer simulation of possible cognitive architectures has shown that serial central processing might be the optimal strategy even if parallel central processing were possible.

Conclusion
The PRP effect refers to a kind of dual-task interference, wherein it is difficult to simultaneously perform two tasks presented in close succession. One of the tasks (usually the one presented second) is performed quite slowly. The PRP effect is remarkably robust, with only rare exceptions, such as when tasks are simple and highly practiced or merely involve movements of the eyes. The primary source of the interference appears to be that central mental processes, such as response selection, must compete for access to limited mental resources. In many cases, there is compelling evidence that central mental operations take place strictly serially—one task at a time. These findings might shed light on the persistent interference that occurs in many real-world situations, such as driving while talking on a cell phone.

Eric Ruthruff
See also Attention, Resource Models; Attention and Action; Automaticity; Divided Attention and Memory; Multitasking and Human Performance

Further Readings


RATIONALITY OF EMOTION

It is sometimes said that emotions in general are irrational or that they cannot be judged in terms of reason: that they are somehow beyond the reach of reason. Yet in particular cases we often deem someone irrational for feeling some specific emotion: “Your anger is unreasonable,” we might say, or “You should be glad that your friend got the job.” The grounds for such judgments, however, remain disputed. Rationality has been exhaustively studied in belief and action. Emotions are causally and conceptually linked to both, but they are not reducible to either belief or action tendencies. If there are canons of emotional rationality as such, they cannot therefore be simply imported from epistemic and practical rationality. This entry explores how far standards of rationality that are derived from the relatively clear cases of belief and action might apply to emotions. Emotions are seen to pose some special problems, particularly in regard to the evaluation of the future or the past and to the elaboration of a relevant concept of consistency. Their rationality is also seen to be significantly affected by social context.

Standard Constraints on Rationality

The criteria of rationality commonly accepted for thought and action suggest four abstract constraints on rationality in general. These provide a starting point for any discussion of emotional rationality.

1. Norms of success. No entity can be assessed for rationality unless it is liable to success and failure. Rationality is not equivalent to success, nor can it ever guarantee success, but X can be said to be more rational than an alternative Y, insofar as X has the greater likelihood of success. Truth is the norm of success for a belief; hence, B1 is more rational than B2 if it is more likely to be true. Similarly, of two alternative actions, the more rational is the more likely to achieve a given goal.

2. Intentionality. The existence of a norm of success implies that whatever can be rational is susceptible of teleological explanation—that is, explanation in terms of some function or purpose. But the converse does not hold. Biological processes typically call for teleological explanations, but only those that are intentional can be said literally to be rational. Intentionality is informally characterized as “aboutness” and is widely thought to be an essential property of mental states. One could speak metaphorically of ants, plants, cells, or even genes as communicating and as choosing alternative strategies of survival or mating. But it would be eccentric to ascribe mentality literally to all biological organisms. Only intentional states can be rational.

3. Origins. Rationality is systematically related to future success, but ascriptions of rationality do not await the verdict of success. On the contrary, rationality hangs on provenance: It depends in part on the origin of the action or belief. If one belief derives from the consultation of astrological signs,
4. **Context dependency.** If origin determines rationality, how do we identify the appropriate antecedents? In statistical reasoning, it may be rational to believe \( p \) relative to one set of facts and \( \text{not-}p \) relative to another, though both sets are equally correct and relevant. Beliefs and actions may be fairly judged both rational and irrational, depending on the extent of the background circumstances taken into account. This context dependency may be illustrated in terms of the tragic real-life case of Andrea Yates, who was induced by voices she heard to drown her five children. At her first trial, the insanity defense was disallowed, on the ground that her careful planning and execution of the drownings proved her rationality. But if one zooms out from its methodical implementation to the project of drowning one’s five children, in obedience to the voice of God, to “save” them, one is bound to see that project itself as irrational. Yet while both Agamemnon and Abraham formed, in obedience to divine command, the project of killing their child, neither is usually thought to have been insane. Unlike truth, a verdict of rationality is never definitive. The framework of its assessment can in principle be extended or modified in an indefinite number of defensible ways.

**Applying Standard Constraints to Emotional Rationality**

To apply these constraints to emotion, we must first make any sense of the notion of emotional “success.” One approach to this is in terms of biological function: Emotions appraise typical life situations and prime the organism for appropriate response; any given emotion is successful if it fulfils that function. This *evolutionary psychological* approach has yielded many insights. But it doesn’t quite get at a notion of success that would be relevant to the question of rationality, for not all functional processes or states are intentional. The intentionality requirement disqualifies even some of the states we loosely refer to as emotions: Moods, insofar as they lack intentionality, also lack conditions of success of the relevant sort. Any emotion that is clearly *about* something, by contrast, intrinsically defines what must be true of its target—the thing, person or situation at which it is directed—for the emotion to be appropriate. That feature is what many philosophers refer to as the emotion’s *formal object*. Truth is the formal object of belief: “Because it’s true” gives a trivial answer to the question “Why do you believe \( p? \)” Similarly, one can give gives a trivial answer to the question “Why are you \( E \)-ing?” for any given emotion \( E \). But there is no global answer for all emotions: Each emotion has its own formal object. Some have obvious names: for fearing, the fearful; being sad, loss; loving, lovable; being disgusted, the disgusting. Others call for awkward explanatory phrases. Just as the formal object of the sense of touch has no single name but relates to hardness, texture, and relative heat, so the formal object of anger has no single name but is awkwardly describable in terms of unjust harm or insult to oneself or others. Whether or not a formal object has a handy name doesn’t seem to correlate with how easily we can tell whether it applies in a particular case. To find something *fearsome*, for example, is to perceive it as *dangerous*: an arguably objective, albeit probabilistic property. By contrast, shame is successful if its object is shameful, but the appropriateness of *shameful*, unlike *dangerous*, does not yield easily to objective confirmation.

**Social Context**

The reason for the difference relates to the third and fourth constraints. While success depends on a match between formal object and target, rationality is determined by an emotion’s origin. If a particular case of fear is caused by factors independent of objective danger, the fear is irrational. Other cases are less clear: There are few objective constraints, for instance, on appropriate origins for love. To be sure, I love him *because he is lovable*, but that places virtually no constraints on the causes of my devotion. Whether he is *really* lovable is a question that barely makes sense, because there is little consensus about the properties that constitute being lovable. Indignation, guilt, and embarrassment, like shame, lie somewhere in between the objectivity of rational fear and the indeterminacy of rational love, but for all of these and many other emotions, what determines rationality are mostly social facts relating to the norms endorsed by members of a given group in historical context.

Where social norms rule, one pertinent dimension of evaluation is *intensity*, a factor generally absent from appraisals of action or belief. In the case of
fear, we have seen that an ascription of irrationality rests in part on the objective absence of danger, but it is also based on an assessment of the intensity of the emotion itself. It seems the intensity of fear should be proportional to the extent of the danger (which in turn consists in both the importance and the probability of the undesirable outcome). The intensity of anger, guilt, or shame is deemed irrational if it is disproportional to the seriousness of the offense inflicted or defect contemplated. Here again, actual social norms and expectations are more important than any other objective facts, and the word rational is interchangeable with reasonable, used to mean something like “I’d feel the same way under the circumstances.”

A good illustration of the interaction of biological origins and social expectations is provided by jealousy, which Jesse Prinz has plausibly characterized as “an acquired blend of anger, fear, sadness and disgust [that] . . . arise in the context of infidelity” (p. 280). If that is right, conditions of appropriateness for jealousy will be highly complex and depend heavily on prevalent social mores in the spheres of sexual orientation and behavior on relations between unmarried persons of the same or opposite sexes, on conventions and expectations governing spouses and sexual exclusiveness, and so forth.

In a different way, the importance of context to assessments of rationality is also illustrated by depth psychology. Recent psychology has confirmed Freud’s observation that our understanding of our own motives is riddled with ignorance and confabulation. The fact that we can be ruled by unconscious motives can turn the rational into the irrational and vice versa. For when we construe an apparently rational act as motivated by an unconscious motive, we expose its irrationality. Conversely, the bearing of an unconscious motive on an acte manqué, or Freudian slip, brings previously inexplicable behavior under the aegis of the belief-desire explanation typical of rational agency.

The Regulation of Emotion

A case frequently discussed is that of recalcitrant emotions, where the judgmental component of an emotion contradicts the agent’s own belief. Such emotions are comparable to visual illusions such as the Mueller-Lyer lines, in that the agent is aware of the failure of fit between target and formal object. It has seemed puzzling, however, that recalcitrant emotions are regarded as irrational, whereas visual illusions are not. Michael Brady has suggested that this is explained by the role of emotions in monopolizing our attention so as to “enhance emotional stimuli.” Where the emotion embodies an evaluation that we know to be incorrect, this is counterproductive. A related but more general explanation is suggested by the long history of disciplines aimed at mastering emotion. From Stoicism through Buddhism and Christianity to René Descartes, David Hume, and William James, it has been assumed that techniques can be devised to correct, redirect, and tame emotion. There is no need, and therefore no comparable set of techniques, for the correction of sensory perception.

Rationality in Memory and Expectation

Worries about emotional rationality range more widely than is accounted for by the framework described so far. What makes for the rationality of emotions toward the past or future is obscure. Is he who regrets always “twice unhappy or twice impotent,” as Benedict de Spinoza (in his Proposition 54) claimed? When evaluating a past experience, should we realistically rate every sequence of moments as the sum of their hedonic value weighted by their duration? Or should we ignore, as Daniel Kahneman has shown we actually do, most of the past moments except for a peak and an end? As for the future, at what rate is it rational to discount it? Sometimes a fervently desired experience proves to be “dust and ashes”; if I predict that I will be disappointed—that I won’t feel as intensely happy when it happens as I now feel I should—should I maximize utility by enjoying the prospect anyway or reduce my anticipation to fit the way I will eventually feel?

Consistency and Constancy

Many puzzles remain concerning both the standards that bear on verdicts of irrationality and the scope of such verdicts. The examples mentioned so far have been of single emotions experienced at some particular time. But one can also ask whether there are constraints on sets of simultaneous emotions and on emotional change through time.

The former question is about emotional consistency. Consistency is not compatibility: If two states are incompatible, they cannot coexist in a single
agent; to say they are inconsistent, by contrast, is to say that they ought not to coexist, which presupposes that they can do so. Whether two emotions are compatible is likely to have to do with the physiological mechanisms that underlie them: If one involves excitation while the other entails inhibition of a certain neurohormonal system, for example, they cannot take place together for mechanical reasons. Genuine inconsistency would derive from the logical conditions on their formal objects. But just what those might be remains obscure.

Questions about the rationality of emotional change through time concern not consistency but constancy. When is it irrational to stop loving someone? How long is it irrational to grieve? The importance of social conventions is particularly obvious in those cases. What remains puzzling, and worthy of investigation, is the extent to which the conventions in question and the variance among them are themselves dependent on biological factors.

Ronald de Sousa

See also Emotion, Cultural Perspectives; Emotion, Structural Approaches; Emotion and Moral Judgment; Emotions and Consciousness; Intentionality of Emotion

Further Readings


REACTION TIME

Reaction time, sometimes referred to as response time or latency, is measured as the time that elapses between the onset of a stimulus and a person’s response to that stimulus. Reaction times (RTs) are widely used in the study of human performance, from testing models of cognitive processing in cognitive psychology to evaluating the design of human-machine interfaces and assisting in diagnoses of such conditions as schizophrenia, learning disorders, and other psychological disorders. This entry presents a brief history of the use of RT, a survey of different kinds of RT experiments, and a summary of how RTs are influenced by other variables.

History

Some of the earliest recorded attempts to evaluate human performance with RT were made by 17th-century astronomers. They worried about the personal equation, which is simply the fact that different observers vary in their estimates of the transit times of stars as the stars moved across the visual field. These astronomers were not as interested in why observers had different personal equations as they were in how much they needed to recalibrate their equipment so that transit times were as accurate as possible. The first serious attempt to explain why RTs varied under different conditions—the first use of RT to determine how people’s brains perform mental tasks—was made by F. C. Donders (1818–1889).

Donders’s idea, called the method of subtraction, was to estimate the time taken by different components of a mental task. The tasks he used are now called simple reactions, go/no go reactions, and choice reactions. These tasks, he reasoned, could be broken down into smaller stages of processing: perceptual encoding, stimulus identification, response selection, and response execution.
Consider, for example, a task where an observer must respond to the presentation of red and green lights by pressing a button. For a simple reaction, an observer presses a button as soon as he sees any light, no matter what color it is. Donders reasoned that this task could be performed only with perceptual encoding and response execution. For a go/no go reaction, however, the observer presses the button only when the light is green. This requires not only the perceptual encoding and response execution stages but also stimulus identification. For a choice reaction, the observer presses one button for a green light and a different button for a red light. The task now requires response selection, by requiring an observer to determine which of the two possible buttons are appropriate for a presented stimulus.

Donders measured his observers’ RTs in the simple, go/no go, and choice reaction tasks. To estimate the duration of the stimulus identification stage, he subtracted the simple RT from the go/no go RT. To estimate the duration of the response selection stage, he subtracted the go/no go RT from the choice RT. This method of subtraction assumed that the task stages were arranged in serial order (so no two stages could be operating at the same time) and that the stages were independent from each other (so if one stage took a very long time to complete, that would not affect how long it took any other stage to complete).

Long after Donders’s seminal work, Saul Sternberg began an investigation of the serial order and independence assumptions. Using a memory search task, in which observers were asked to determine whether a numeral had been previously presented in a small set of numerals, he observed that (choice) RT to respond “yes” or “no” increased as a linear function of the size of the set. He argued for a serial process, in which the target numeral is compared to each numeral in the memory set one at a time. He then extended this logic to task stages in his additive factors method.

The additive factors approach examines the difference between RTs in different experimental conditions. In particular, the method requires that the experimenter identify experimental factors that selectively influence different independent stages of processing. For example, the memory search task presumably requires at least two stages: perceptual encoding and memory search. Making the stimuli difficult to see will prolong the encoding stage but should not influence the search stage. Similarly, increasing the set size should prolong the search stage but not the encoding stage. A factorial design should, therefore, produce additive effects of encoding difficulty and search set size—that is, there should be no interaction between these factors—if perceptual encoding and memory search are serial processes.

Sternberg’s additive factors method is still very influential, even though it can be demonstrated that other kinds of mental architectures can produce RTs that are indistinguishable from those produced by a serial process. Donders’s and Sternberg’s methods use measures of mean RT, and many researchers still rely on mean RT to explore hypotheses about mental processing. However, many researchers now work with models of the cognitive system that can predict the entire distribution of RTs, as well as the accuracy of different responses. These models have led to the development of new experimental designs as well as new methods of analysis.

RT Experiments

RT experiments can be categorized according to the extent to which information provided by the stimulus is compressed in the response. Many RT experiments have fewer possible responses than the number of stimuli that can be presented. Simple RT tasks can present any number of different stimuli, but only a single response is required to all of them. Choice RT tasks have more than a single response but usually fewer possible responses than possible stimuli. Two-choice RT tasks are by far the most common, in which observers are asked to respond “yes” or “no” to a potentially very large number of stimuli. Word recognition memory experiments, for example, ask observers to respond “yes” (old) to studied words and “no” (new) to novel words. The number of studied and novel words can number in the hundreds, but the number of responses is only two.

Choice RT tasks may have more than two possible responses, in which case they are called N-choice RT tasks. For example, a word recognition memory experiment may ask observers to rate their confidence that a word was previously studied or not by using an N-point scale (e.g., 1 = confident old, 2 = probably old, 3 = maybe old, . . . 6 = confident new). Such a task is sometimes called a judgment task. Categorization tasks, where objects are to be classified as members of N different groups, are also N-choice RT tasks. When the number of
possible responses is equal to the number of possible stimuli, the choice-RT task becomes an identification task—each stimulus is identified by one and only one correct response.

The go/no go task can be viewed as a choice-RT task where one of the possible responses is not to respond at all. Related to the go/no go task are stop-signal tasks. These are choice-RT tasks in which, for some trials, a stop signal is presented at some time after the stimulus, indicating that the observer should not respond. Stop-signal tasks are used to explore the dynamics of response preparation. As the stop-signal delay increases, observers are less able to inhibit their responses, which suggests that the choice process has components that are gradual and build up over time.

The stop-signal task asks observers to do two things at the same time: Select a response to a stimulus and also prepare to stop that response. Two other kinds of RT experiments that are closely related to the stop-signal task are dual-task and task-switching experiments. These are used to explore mechanisms of response inhibition and automaticity of processing. Results from these kinds of experiments inform our understanding of executive processing, or how people are able to control their behaviors, starting and stopping them at appropriate times, and also of the factors that contribute to uncontrollable (automatic) behaviors.

Influences on RT
Apart from any independent variables an experimenter chooses to manipulate in an experiment, there are a number of other variables that influence RT in simple, choice, and identification tasks. If these other variables are not controlled, they may confound the independent variables. Some variables that influence RT are arousal, age, intelligence, and fatigue. Other variables are more specific to the kind of task the observer is asked to perform.

In simple-RT tasks, RT is positively related to stimulus intensity (e.g., brightness or loudness); as stimuli become easier to detect, simple responses to them can be made more quickly. This relationship is captured by Piéron’s law, which states that mean RT is equal to \(a + bI^c\), where \(a\), \(b\), and \(c\) are parameters, all greater than zero, to be estimated from the data, and \(I\) is stimulus intensity. Modality is also very important in simple-RT tasks: Auditory stimuli evoke faster RTs than visual stimuli, for example.

Choice-RT will increase as both the number of stimuli and responses increases. This is the Hick-Hyman law of mean RT: RT is a linear function of the amount of uncertainty in the task. Uncertainty is a dimensionless quantity that depends on the number of possible outcomes in an experiment and their probability. It can be used to describe many things, but in this context it refers to the amount of information provided by the occurrence of an event. For example, if there is only one possible response, observing that response does not tell you very much. If there are \(n\) equally likely responses, observing one of them is very informative.

Both choice-RT and identification tasks may be influenced by stimulus-response compatibility. Compatibility refers to the ease with which stimuli can be associated to different responses. Compatibility experiments have often focused on the spatial features of stimuli and responses or on where the stimuli appear relative to the location of the responses to be made to them. Highly compatible spatial relationships (e.g., responding with a right button to the stimulus that appears on the right) result in faster RTs than less compatible spatial relationships. Compatibility effects can also occur when the spatial stimulus dimension is task irrelevant and for nonspatial stimulus dimensions such as positive-negative affect of stimuli and verbal responses.

Perhaps the most important issue that arises in choice-RT experiments is the correlation between accuracy and RT: The faster an observer responds, the more errors she makes. Earlier studies such as those of Sternberg and Donders assumed that if the error rate were small then error responses could be safely ignored. While in general this is true, it is also true that very small changes in error rate may reflect a very large change in processing strategy and hence RT.

For this reason, there has been much attention paid to the speed-accuracy trade-off in RT experiments. Much of our current understanding about how people make simple decisions has come from models of RT designed to explain the speed-accuracy trade-off. These models assume that responses arise from a process of information accumulation. Observers sample evidence toward alternative responses from a display, and when that evidence reaches a threshold, a response can be initiated. The speed-accuracy trade-off arises naturally as subjects raise and lower their thresholds for responding. If thresholds are low, it will require less evidence and
therefore less time to reach them, but it will be easier for an inappropriate response to accumulate evidence enough to reach a lower threshold. If the thresholds are higher, RTs will be slower, but inappropriate responses will be less likely to reach the threshold.

Evidence accumulation models, sometimes called sequential sampling models, now are applied to problems across the spectrum of brain sciences, including neuroscience aging, developmental disorders, and clinical psychology. We now have good evidence suggesting that at least some parts of the brain function control decisions through groups of neurons that act as evidence accumulators.

Conclusion

It has always seemed incredible so simple a dependent variable as RT could ever tell us anything really important about how the brain works. As R. Duncan Luce described in 1986, this endeavor is similar to trying to reverse engineer the motherboard of a computer by measuring how long it takes to run different programs. In fact, using RT alone, it will not be possible to learn anything of the brain’s intricacies. However, by linking measurements of RT to other behavioral variables and to neural data such as single-cell recordings or functional magnetic resonance imaging, we have learned a great deal about how simple choices are made, including how people control their behaviors and attend to different features of their environments. Understanding these simple choices is fundamental to understanding more complex human behavior.

Trisha Van Zandt

See also Aging, Memory, and Information Processing Speed; Automaticity; Decision Making and Reward, Computational Perspectives; Divided Attention and Memory; Multitasking and Human Performance; Psychological Refractory Period; Stroop Effect

Further Readings

Luce, R. D. (1986). Response times: Their role in inferring elementary mental organization. New York, NY: Oxford University Press.


REALISM AND INSTRUMENTALISM

The choice between realism and instrumentalism is at the core of concerns about how our scientific models relate to reality: Do our models aim to be literally true descriptions of reality, or is their role only to be useful instruments for generating predictions? Realism about X, roughly speaking, is the claim that X exists and has its nature independent of our interests, attitudes, and beliefs. An instrumentalist about X denies this. He or she claims that talk of X should be understood as no more than a locution for generating predictions; such talk should not be understood as taking on a commitment to the existence of X. According to an instrumentalist, we should either flatly not believe that X is out there or else should suspend judgment about the existence of X. The most we need acknowledge is that talk of X is useful in making predictions.

The question of realism versus instrumentalism can be asked about almost any theoretical entity in science. It is likely, and seems reasonable, that different answers will be given in different cases. Someone may wish to be a realist about certain theoretical entities (e.g., electrons) but an instrumentalist about others (e.g., centers of gravity). Not every noun phrase in a scientific theory should be taken as expressing...
an ontological commitment. Psychological theories are no exception. Almost every theoretical entity posited by psychology has been questioned as to whether it is really out there or just a useful theoretical fiction. This entry focuses on two such major theoretical entities: (a) propositional attitudes (e.g., beliefs, desires) and (b) conscious states (qualia).

Propositional Attitudes
Psychological theories, both in science and our folk conceptions, often use propositional attitudes (beliefs, desires, hopes, assumptions, fears, etc.) to explain and predict how people think. These mental representations seem to figure as causal agents explaining and predicting how people think. These mental representations are the causal agents behind a computer's behavior. Just as patterns of electrical activity controlling a computer's machine code control a computer's behavior and cause the occurrence of further expressions of machine code inside the machine, so sentences in our language of thought enter into causal relations, control our behavior, and cause the occurrence of new sentences in our language of thought (new beliefs, desires, etc.). Like a computer's machine code, sentences in our language of thought exist as a pattern of physical activity inside our heads. Hence, propositional attitudes such as beliefs and desires are discrete, reoccurring, entities with causal powers. They are the causal agents behind our behavior, just as patterns of electrical activity instantiating a computer's machine code instructions are the causal agents behind a computer's behavior.

In contrast, Paul Churchland argues for a robust form of instrumentalism about propositional attitudes. According to Churchland, belief and desire terms fail to latch onto any entities in the world and at best serve as a useful way of talking for ordinary folk. Churchland concedes that psychological theories employing propositional attitudes enjoy some predictive success; however, he thinks that Fodor overestimates the degree of that success in light of the potential of future neuroscientific theories to explain behavior without reliance on propositional attitudes. Churchland points out that many theories are instrumentally useful yet false. Ptolemaic astronomy, which posited celestial spheres, makes many true predictions but is nevertheless false. Churchland claims that beliefs and desires will go the way of celestial spheres. Churchland's main argument for the nonexistence of propositional attitudes can be broken into two steps. The first step is to argue that folk psychology is a theory and that the meaning of propositional attitude terms (expressions such as belief and desire) is fixed by their role in that theory. What belief and desire mean is wholly, and exclusively, specified by folk psychology: What it means to be a desire is to be something that combines with beliefs and causes action in precisely the way described by folk psychology. Churchland's second step is to argue that folk psychology is false. If the folk psychology theory is false, then nothing satisfies the role ascribed to beliefs and desires, and consequently, beliefs and desires, as traditionally conceived, do not exist. Both steps in Churchland's argument have been questioned. Against the first, Ronald Mallon and colleagues have argued that it is far from clear the extent to which the meaning
of propositional attitude terms such as belief and desire ride on the fortunes of folk psychology. It is not obvious how much, or indeed if any, of folk psychology needs to be true in order to fix the meanings of propositional attitude terms. Against the second step, Terence Horgan and James Woodward have argued that Churchland underestimates the success of folk psychology and overestimates the demands we should place in order to be justified in believing it is true.

There are many ways of developing the instrumentalist thought. Daniel Dennett offers a milder form of instrumentalism about propositional attitudes than Churchland. On Dennett’s view (unlike Churchland’s), talk of beliefs and desires is true, but (unlike Fodor’s) such talk does not succeed in referring to entities that have an objective existence or representational content independent of our interests. According to Dennett, what is involved in an agent having a belief or desire is not that there is a discrete entity inside the agent—the belief that p—with causal powers pulling the strings behind the agent’s behavior but merely that there is a predictable payoff in describing the agent as if it were controlled by such an entity. To describe an agent in terms of propositional attitudes is to adopt what Dennett calls the “intentional stance”: a mode of explanation that attributes to the agent the beliefs and desires that a rational being placed in its shoes ought to have. According to Dennett, if the intentional stance is reliable as a method of predicting the behavior of a system S, then ipso facto system S has those beliefs and desires. All that is required for a system to have a belief is for it to be useful in predicting the behavior of that system to assume that it has that belief.

A consequence of Dennett’s instrumentalism is that beliefs are easy to achieve. It is often helpful to predict the behavior of cats, robots, washing machines, computers, plants, bacteria, cars, and thermostats by treating them as if they have beliefs and desires. According to Dennett, there is no difference between this as if and genuine possession of beliefs. It also becomes harder for some systems to achieve beliefs on Dennett’s view. Patients suffering from mental illness often cannot be profitably viewed as rational agents when it comes to predicting their behavior. Therefore, on Dennett’s view those patients lack beliefs and desires. Their behavior would have to be explained in some other way than intentional psychology—for example, by dropping down to the level of their neurophysiology.

Dennett’s instrumentalism about propositional attitudes raises questions about exactly how one should draw the line between realism and instrumentalism. The simple characterization of realism and instrumentalism given at the beginning of this entry fragments into a number of different theses that can, in principle, be affirmed or denied separately. Dennett denies two key realist theses about propositional attitudes: (a) Mind-independence—propositional attitudes of an agent have their existence and nature independent of the interpreting interests of observers; (b) discrete causal powers—propositional attitudes are discrete, reoccurring, entities inside the head with the causal powers to produce behavior. However, in contrast to Churchland, Dennett affirms a realist intuition about propositional attitudes: (c) existence—propositional attitudes exist—they are really out there, unlike celestial spheres. According to Dennett, propositional attitudes exist as patterns that are available to an interpreter to be used for prediction. These patterns are “real” and “objective” in the sense that there are objective facts about what is and what is not predictively successful to assume within the intentional stance. In other words, some belief and desire attributions pay off in that they yield successful predictions of behavior, and others do not. These real patterns of predictive success and failure are the facts in the world that make claims about propositional attitudes true or false.

**Qualia**

We often report that there is a phenomenal aspect to our experience: Seeing red feels a certain way, having a mouse cupped in one’s hand feels a certain way, and the qualitative aspects of different experiences are different. Distinctive qualitative experiences accompany large parts of our mental life. Do our reports of qualitative experiences describe really existing entities (phenomenal properties or states inside our head), or do they serve some other purpose? Are qualia real, or does talk of qualia fail to refer to anything in the world?

An immediate problem that realism and instrumentalism about qualia face is that it is hard even to describe what qualia are, and hence, hard to say what one is or is not being a realist about. Often, the best one can do is point to examples of qualia, such as those above. Daniel Dennett develops a strong instrumentalist line against qualia. His target is the widespread assumption that qualia have certain
special properties: They are ineffable, intrinsic properties of experience, private and directly accessible to the experiencer. Dennett argues that nothing satisfies this specification, and hence, there are no such things as qualia. Dennett’s position is similar to Churchland’s strong line against propositional attitudes: Just as the falsity of Ptolemaic astronomy justifies the inference that there are no celestial spheres, so the falsity of philosophical claims about qualia justifies the inference that there are no qualia. Talk of qualia still serves a purpose according to Dennett in that it provides a shorthand summarizing our ability to detect certain properties in the world, such as color properties, that lack a compact description in any other terms.

Dennett’s instrumentalism has drawn heavy criticism, not least because it runs up against the robust impression that there are real qualitative properties of our mental states that have at least some of the properties mentioned above. However, realism comes at a price. If one grants realism about qualia, then the question arises of what kinds of entities qualia are. Are qualia represented features of the world encoded by our nervous system, similar to our unconscious encoding of features of the world such as position, size, and shape information? If so, what makes conscious “felt” representations different from unconscious representations? Or are qualia intrinsic physical properties of our nervous system, independent of our ability to represent? Or are qualia something different entirely, requiring properties that float free from the physical world and any representations it encodes? No consensus currently exists to the question: the hard question of consciousness—the nature of qualia under a realist understanding.

Mark Sprevak

**See also** Access Consciousness; Eliminative Materialism; Folk Psychology; Mind-Body Problem; Naïve Realism; Representational Theory of Mind; Smell, Philosophical Perspectives

**Further Readings**


**Reductive Physicalism**

One way in which something utterly mysterious can be unraveled is to reduce it to something we can understand. Not surprisingly, then, philosophy, which often tries to understand mysteries, takes reduction as one of its tools. Is knowledge reducible to true, justified belief? Are moral codes reducible to personal preferences? Are meaningful statements reducible to constructions of immediate experience? And is the mind reducible to the brain? This entry provides an introduction to the concept of reduction, as it is used in philosophy of mind, and an overview of the various reductive accounts of the mind posited by philosophers, culminating with a brief discussion of the relative merit of these views and of reduction in general.

**Reduction**

What is reduction? What does it mean to say that the mind is or is not reducible to the brain? Like many terms, both in and out of philosophy, the word *reduction* is used in a variety of ways. In everyday language, we often understand reduction, quite naturally, to mean lessening or reducing in size. You might go on a reducing diet—for example, by eliminating those luscious French sauces that require reduction when cooking. In philosophy, the word reduction is sometimes used with this ordinary meaning, as philosophers often take it to mean simplification. When we reduce one ontological category to another, by showing the one category
to be nothing but the other, we simplify our ontology by cutting down the number of kinds of things that exist in the world. However, the philosophical notion of reduction can also encompass an explanatory element. One category reduces to another, in this sense, if can be explained in terms of the other.

In philosophy of mind, these two notions of reduction—the *nothing but* notion and the *explanatory* notion—often (though, as we’ll see, not always) overlap. For example, one way in which the mind can be understood as nothing but the brain (or, more accurately, certain parts of the brain) is by reductively explaining mental processes entirely in terms of the neural processes. Not all explanations of one thing in terms of something else are reductive explanations. I might explain the reason for the spill on the floor in terms of why it happened: My cat jumped on the table. But reductive explanations are not like this, they do not tell us the cause of something; rather, they aim to tell us what something is. For example, genetic material is reductively explained in terms of DNA, water in terms of $\text{H}_2\text{O}$, temperature in a gas in terms of mean molecular momentum, and if the mental is reductively explainable in terms of neural processes, mental processes in terms of neural ones. For example, one version of reductive physicalism holds that pain is reducible to C-fiber activity (or, more accurately, to whatever neural process that is found to be perfectly correlated with pain). Such a reduction aims to explain pain by showing that it is nothing but C-fiber activity. If pain could be entirely explained in terms of processes in the brain, such as C-fiber activity, we would have both simplified our ontology and found out what pain really is.

**Physicalism**

What makes reductive physicalism a form of physicalism? Physicalism holds, roughly, that everything is physical. Reductive physicalism is simply the view that the mind is reducible—in one of the various senses to be specified below—to either neural processes or behavior. Assuming that neural processes and behavior are both entirely physical, such a view deems physicalism’s primary nemesis, the mental, to be physical.

**Eliminative Physicalism**

The explanatory and simplifying elements of reduction sometimes come apart. For example, some see neuroscience as paving the way toward an “eliminative reduction” of at least some of what we currently think of as mental processes. Those who uphold such a view, eliminativists, as they are called, think that we should not expect to explain the mental in terms of any other more fundamental theory. But this is not because they hold, as does the dualist, that mental processes are distinct from physical ones. Rather, they hold that much of what we currently think of as mental is just a made-up fiction. Correlations between the mental and neural, they believe, will not be found. Thus the mind, they think, should go the way of witches and phlogiston. Although it was once commonly thought that witches existed, we now simply deny their existence; although it was once thought that when wood burns or metal rusts phlogiston is released, we now deny that phlogiston exists. The mind, according to eliminativists, is more like phlogiston than like the gene. Eliminativism is reductive in as much as it simplifies our ontology (reduces it in the ordinary sense of the term), but rather than explaining the mental in neurological terms, it explains it away.

Eliminativism, apart from being reductive in one sense, is also most definitely a case of physicalism, since it posits only physical processes. Although dualism is not a reductive theory, other antiphysicalist positions can be thought of as reductive. For example, phenomenalists hold that physical objects are nothing but sense data; the tomato on your plate, on this view, is something like a concatenation of mental images. Idealists are also reductivists of an antiphysicalistic stripe. For the idealist, physical objects are reducible to ideas. The reductive physicalist, however, turns this theory on its head: Ideas (and everything else mental) are reducible to physical objects.

Eliminative reduction is a version of reductive physicalism, but it is not the most common one. Many physicalists are optimistic that we will find robust correlations between mental and physical processes. And among these some accept that what we think of as mental can, at least in principle, be explained by physical processes, while others think that although mental processes, such as twinges of pain and showers of tickles, will never be explained in terms of physical processes, mental processes are identical to physical processes nonetheless. These two groups of reductive physicalists do not advocate eliminative reduction, as they maintain that all or nearly all that we know and love about the
mental exists. Rather, they advocate what is sometimes called retentive reduction; eliminative reduction eliminates the mental from our ontology, while retentive reduction involves explaining it or at least identifying it in terms of the brain.

**Explanatory Mind-Brain Identity Theory**

One form of retentive reduction in philosophy of mind is the mind-brain identity theory, also called the type identity theory or sometimes just the identity theory. Philosophers who accept the identity theory think that mental processes are one and the same thing as neural states. For example, the type of thing we call pain, according to this view, is nothing more than the type of thing we call C-fiber stimulation (or, rather, the type of thing we call \( P \), where \( P \) refers to the neural state that is perfectly correlated with experiences of pain). Pain exists, claims the identity theorist, and it is reducible to activity in the brain. Just as the uncanny connection between Dr. Jekyll and Mr. Hyde led the London lawyer Utterson to eventually conclude that the misanthropic Hyde was none other than his dear old friend Dr. Jekyll, those who uphold the identity theory think that a correlation between pain and C-fiber stimulation might lead us to conclude that pain and C-fiber stimulation are one and the same thing as well.

**Mysterianism**

As some see it, the mind-brain identity theory is explanatory in the sense that it provides what some see as a simple answer to the question of why pain is correlated with C-fiber stimulation: The correlation holds because pain is C-fiber stimulation. If pain is just C-fiber stimulation, if *pain* and *C-fiber stimulation* are just two names for the same thing, there is no more question about why every time people are in pain their C-fibers are firing than there is a question about why every time I shut the door I also close it.

Other reductive physicalists, however, although they accept the identities, deny that identities of this form provide explanations; they still think that there is a question of how the activity of C-fibers could be pain or, to use another example, how the activity of pyramidal cells could be consciousness. Pain, in their view, is reducible to neural processes in the sense of it being nothing but something going on in the brain, but as they see it, we lack an explanation of how this could be so.

**Behaviorism**

So we have two different sorts of reductive accounts of the mind: eliminativism and the type identity theory. And the latter can be divided into explanatory type identity and what is sometimes called mysterianism. Another form of reductive physicalism is behaviorism. On the most basic understanding of this view, the mind is nothing over and above behavior. Pain, for example, is just the bodily movements you make when you are in pain. The desire for carrots is just the bodily movements you make when you desire carrots. An obvious problem with this view is that sometimes one deliberately suppresses such movements. If it’s not mealtime, for example, you’ll need to suppress your urge to open the fridge. However, more sophisticated forms of behaviorism claim not to reduce pain to certain forms of behavior, per se, but rather to the *disposition* to behave in certain ways. To be fragile, for example, is to have the disposition to break if dropped on a hard surface. A glass has this disposition even when it is not dropped. Similarly, the behaviorist will say, to be in pain is to have the disposition to wince, say “ouch,” and so forth.

**Cost-Benefit Analysis**

Are there reasons to accept one reductive view over the others? All have at least some apparent flaws. Eliminativism is criticized for being too pessimistic about the prospects for mind-brain correlations. Neuroscience, the argument goes, is in its infancy and it may be that what seems utterly inexplicable neuroscientifically now will be understood later. And behaviorists are criticized for not being able to spell out the relevant dispositions. What, exactly, are the relevant background conditions under which you would say that you are in pain? What, exactly, are the relevant mitigating circumstances? It is not at all clear how to answer these questions. Moreover, there are some mental states, such as the belief that seven is a prime number, that do not even have any standard associated behaviors. What does one typically do when one believes that seven is a prime number? The question is rather absurd.

The mind-brain identity theory, according to which mental process \( M \) is necessarily physical
Regret

We feel regret when realizing or imagining that our present situation would have been better had we decided differently. This entry addresses how regret is experienced, whether we regret actions more than inactions, how regret relates to counterfactual thinking and decision making, and how regret influences behavior.

The Experience of Regret

Regret stems from a comparison between outcomes of chosen and nonchosen alternatives in which the latter outperform the former. It is a painful emotion that reflects on our own causal role in the current,
suboptimal situation. The emotion regret is accompanied by feelings that one should have known better and having a sinking feeling, by thoughts about the mistake one has made and the opportunities lost, by tendencies to kick oneself and to correct one’s mistake, by desires to undo the event and get a second chance, and by actually doing this if given the opportunity. It is a cognitively based emotion that motivates one to think about how the current negative event came about and how one could change it or how one could prevent its future occurrence.

**Action and Inaction Regret**

We may regret sins of omission and sins of commission. Early regret research indicated that people tend to regret their actions (commissions) more than their inactions (omissions). Later research showed that time plays a crucial role here. In the short run, people tend to feel more regret over their actions (the stupid things they did or bought), but in the long run, they tend to feel more regret over their inactions (the school they never finished, the career or romance never pursued). This temporal pattern to regret is due to a number of factors that decrease the regret for action over time (e.g., we take more reparative action and engage in more psychological repair work for action regrets than for inaction regrets) and to factors that increase the regret for inaction over time (e.g., over time we may forget why we did not act on opportunities, making the inaction inexplicable).

Another factor determining the intensity of regret is the justifiability of the decision. People feel most regret over decisions that are difficult to justify. Decisions that are based on solid reasons produce less regret than decisions that are not well thought through. This justifiability may also explain when actions are more regretted than inactions and when the reverse is true.

**Regret, Decision Making and Counterfactual Thinking**

Regret is unique in its relation to decision making. One only experiences regret over a bad outcome when at some point in time one could have prevented the outcome from happening. Other emotions can also be the result of decisions; for example, one may be disappointed with a decision outcome or happy about the process by which one made a choice. But these emotions can also be experienced in nonchoice situations. For example, one can be disappointed in the weather and happy with a birthday present, but one cannot regret these instances (unless, of course, if the disappointing present was suggested by oneself).

The relation between regret and decision making is also apparent in regret’s connection to counterfactual thinking. Counterfactual thoughts are thoughts about what might have been. Note that not all counterfactual thoughts produce regret but only those that change a bad outcome into a good one by changing a prior choice or decision. Thus, when it rains on the way home from work and I get wet, I feel regret when I generate a counterfactual thought in which I brought an umbrella but not when I generate a counterfactual in which it would be a beautiful day. In the latter case, counterfactual thoughts about better weather that could have been would result in disappointment but not in regret (I could not change the weather, so there is nothing to regret).

**Experienced and Anticipated Regret**

Psychologists became interested in studying regret partly because it is not only a passive emotional reaction to bad decisions but also a major influence on our day-to-day decision making. This influence can take two forms. First, the experience of regret may produce a behavioral inclination to reverse one’s decision or undo the consequences. Second, decision makers may anticipate possible future regret when making decisions and choose in such a way that this future regret will be minimal.

The influence of retrospective regret on behavior can be functional. The aversive experience prompts us to undo the cause of the regret. For example, after buying a product that proves to be suboptimal, regret can motivate us to ask for our money back, or it may result in apologies in the case of interpersonal regrets. In both instances, regret can help us to satisfy our needs. It protects us from wasting money and helps us maintain good social relationships. Additionally, regret can be functional in the sense that the painful self-reflective nature of the experience is one of various ways by which we learn. The feeling of regret over bad decisions and wrong choices makes them stand out in our memory and helps us make better decisions in the future.

The idea that people, when making decisions, might take into account future emotional reactions
Rehearsal and Memory

Rehearsal refers to the overt or subvocal recitation of to-be-remembered (TBR) verbal material during encoding, storage, or retrieval from memory. This entry reviews the considerable evidence that rehearsal is a major determinant of memory performance in many situations, involving both short-term and long-term retention. Rehearsal has played an explanatory role in otherwise highly diverse theories of memory, for example that of Alan Baddeley. In recognition of the potential circularity of an unobservable explanatory construct (e.g., whenever memory performance is good, there must have been effective rehearsal, and whenever performance is poor, rehearsal must have been absent), much effort has focused on the identification, operationalization, and experimental control of rehearsal.

Elaborative Versus Maintenance Rehearsal

One influential proposal involved the distinction between elaborative rehearsal (also known as Type II rehearsal) and maintenance (Type I) rehearsal advanced by Fergus Craik: During elaborative rehearsal, new material is meaningfully related to other information—for example, by deciding whether a TBR word fits into a sentence. Maintenance rehearsal, by contrast, involves rote repetition of items without relational processing, usually with the express purpose to maintain already encoded information in awareness.

There is consensus that increasing elaboration of TBR material leads to better long-term retention; for example, judgments involving the meaning of the TBR material lead to better subsequent recall than judgments of its sound, which in turn yields better memory than judgments of surface structure (e.g., reporting the number of letters). By contrast, the effect of maintenance rehearsal is more ambiguous:

See also Decision Making, Neural Underpinnings; Decision Theory, Philosophical Perspectives; Emotion Regulation; Rationality of Emotion

Further Readings


On the one hand, when people are led to believe that long-term retention of the material is unimportant, the number of overt recitations of TBR words is often unrelated to subsequent surprise recall performance. On the other hand, when memory is tested by recognition, the duration of maintenance rehearsal is usually correlated with performance, as shown by Robert Greene in an extensive review. The distinction between maintenance and elaborative rehearsal might therefore be best considered as a *continuum* involving the amount of attention that is paid to rehearsal. When attention is minimal, then the number of rehearsals may fail to correlate with measures of memory. In all other situations, the number of articulations of TBR material is correlated with enhanced performance.

**Linking Rehearsal to Performance**

Going beyond mere correlation, support for a causal link between rehearsal patterns and memory performance has been adduced both empirically and theoretically. Lydia Tan and Geoff Ward provided a detailed analysis of rehearse-out-loud protocols in free-recall experiments. In immediate free recall, performance at all list positions was primarily a function of the recency of an item’s last rehearsal, to the exclusion of other experimental variables (e.g., presentation rate, list length, and participants’ age) that affect recall when rehearsal patterns are ignored. It follows that those other variables exert their effect on performance only indirectly, via modulation of people’s rehearsal patterns. It is noteworthy that the recency-of-rehearsal analysis accommodated the primacy effect (better recall for items that were presented early in the list), because early items tended to be rehearsed again much later in the list. In delayed recall, other factors such as the number and spacing of rehearsals become relatively more important than recency, and those factors may also explain why maintenance rehearsal is sometimes ineffective.

Intriguingly, Tan and Ward found no additional advantage for spontaneous self-generation of the rehearsal schedule in comparison to articulation of an identical protocol that was generated by a different participant. The process of choosing items for rehearsal thus appears less important than their effective re-presentation by articulation.

David Laming presented a detailed mathematical model that predicts entire sequences of recall from the pattern of overt rehearsals and item presentations during study. Laming’s theory emphasizes the importance of considering subject-generated events (rehearsal) on par with experimenter-controlled events (presentation of study items). Overall, rehearsal patterns indubitably are a major determinant of long-term memory performance, and the success of recall can be predicted from preceding rehearsals with quantitative precision. To a first approximation, the data warrant the conclusion that an item’s probability of recall is determined by the recency of its last rehearsal.

**Rehearsal in Memory Over the Short Term**

Turning to memory over the short term, relevant models have frequently invoked rehearsal as an explanatory construct. In contrast to long-term memory research, the imputed role of rehearsal in short-term memory has focused less on encoding and more on the protection of information against loss, often presumed to occur through temporal decay—that is, the inexorable decline in memory strength over the passage of time.

Accordingly, research in short-term memory has often focused on the effects of wholesale manipulations that prevent rehearsal altogether. Rehearsal is commonly thought to be abolished by *articulatory suppression* (AS from here on)—that is, the repeated articulation of an irrelevant word. AS is particularly diagnostic because, if it successfully suppresses rehearsal, then the effects of temporal decay should become observable. Contrary to that expectation of decay-based models, Stephen Lewandowsky and colleagues have shown in 2004 and 2008 that performance is unaffected when AS-filled retention intervals are extended. The hitherto tight coupling between a presumed mechanism of forgetting—namely, decay—and rehearsal therefore appears unnecessary.

Tan and Ward in 2008 reported the only analysis to date of (spontaneous) overt rehearsal protocols involving short lists and immediate serial recall (i.e., the recall of list items immediately upon presentation in their original order). Performance was found to be strongly associated with the length of the longest forward-rehearsal sequence; that is, a person who rehearsed the first five items in sequence performed better than a person who rehearsed only two items in sequence.
Further evidence for the role of rehearsal over the short term has evolved from research in working memory. A popular working-memory paradigm interleaves the TBR material with unrelated processing stimuli (e.g., by presenting arithmetic tasks in between list items). Recent models of this “complex-span” task proposed by Pierre Barrouillet and colleagues in 2004 have again coupled rehearsal with decay as two opposing processes. This coupling is based on the finding that memory performance is a function of the proportion of time—rather than its absolute duration—in between TBR items that is taken up by the processing task. Memory is presumed to decay during the processing but is restored during brief intervening pauses by a rehearsal process. Although the work by Lewandowsky and colleagues provides reasons to question the presence of decay, there is little doubt that even brief pauses in between processing episodes can be used to restore memory traces; notably, this form of rehearsal appears to be attentional rather than articulatory.

Direct behavioral evidence for the presence of attentional refreshing consists of the fact that it can occur concurrently with overt recitation of verbal material. It is known that different brain regions are involved in attentional and articulatory rehearsal and that the two forms of rehearsal make separate but additive contributions in a complex-span task.

Stephan Lewandowsky

See also Memory Recall, Dynamics; Working Memory

Further Readings


REINFORCEMENT LEARNING, PSYCHOLOGICAL PERSPECTIVES

Reinforcement is the process by which experience changes both overt behavior and that more subtle behavior called cognition. The ability of individual experience to change behavior is arguably the most important contribution of natural selection by the ancestral environment. Whereas natural selection adapts the species of which the individual is a member to the more constant contingencies of the ancestral environment, selection by reinforcement adapts the individual to the variable contingencies of his or her personal environment. In natural selection, the environment selects for structural and functional characteristics that favor reproductive fitness—the survival of offspring. In selection by reinforcement, the environment—including interactions with others—selects behavior that adapts organisms to the specific events encountered in their lives.

The entry summarizes experimental research that identifies the factors that select for changes in individual behavior. The entry then indicates some of the neural mechanisms that implement selection by reinforcement. A core belief of many learning theorists is that reinforcement will ultimately play a central role in understanding the origins of complex behavior that is analogous to the role played by natural selection in understanding the evolution of complex species. That is, the cumulative effects of selection by reinforcement will prove competent to produce the complexity and diversity of individual behavior. The entry concludes with a discussion of methods to explore the implications of reinforcement for complex behavior.

Behavioral Analysis of Reinforcement

Experimental research on the factors necessary for selection by reinforcement began around the turn of the previous century in the laboratories of Ivan Pavlov in Russia and Edward Thorndike in the United States. Both sought to control their experiments through the use of nonhuman animals in the belief that natural selection would exploit largely common learning processes that exist in a wide range of species. In Pavlov’s procedure, called classical conditioning, a neutral stimulus (such as the
sound of a tone) occurs before another stimulus (such as taste stimulated by food in the mouth) that already evokes behavior (such as salivation). After several occurrences of this sequence of events, the formerly neutral stimulus begins to evoke control of the salivary response. The stimulus that evoked behavior at the beginning of the procedure (food in the example) functioned as a reinforcing stimulus, or reinforcer. Reinforcers change and strengthen the ability of stimuli to guide behavior. Pavlov's procedure enabled an arbitrary stimulus (the tone) to guide a response (salivation) that was originally evoked by another stimulus (food). In Thorndike's procedure, called operant conditioning, a response (such as pressing a lever) occurs before a stimulus (such as food) that already evokes behavior (such as salivation). After several such sequences of events, lever pressing and salivating both become more likely when the learner sees the lever. Thus, the operant procedure enables the environment to guide two sets of responses—the relatively arbitrary response that preceded the reinforcing stimulus (lever pressing) as well as the response evoked by the reinforcer (salivating). B. F. Skinner most clearly appreciated that the operant procedure allowed the full behavioral capabilities of the learner to be modified by experience, not just those responses that could already be elicited by some stimulus. As experience accumulated—thus expanding the behavioral repertoire of the learner—the potential increased for the selection by reinforcement of even more complex behavior. The change in the guidance of behavior produced by both the classical and operant procedures is called conditioning because the change is conditional on (that is, dependent on) events in the individual experience—a stimulus-reinforcer sequence in Pavlov's case and a response-reinforcer sequence in Thorndike's case.

Research with both the classical and operant procedures indicates that the events in the sequence must occur very close together in time for conditioning to occur. In the classical procedure, the neutral stimulus must precede the reinforcing stimulus by no more than a few seconds if the neutral stimulus is to acquire control of the reinforcer-elicited response. Similarly, in the operant procedure the arbitrary response must immediately precede the reinforcing stimulus if the environment is to acquire control of the arbitrary response as well as the reinforcer-elicited response. Together, findings from these procedures identified the first factor that is necessary for selection by reinforcement—temporal contiguity between the events.

The second factor necessary for selection by reinforcement was discovered through an elaboration of the basic classical procedure: A response was first conditioned to one stimulus (S1) by presenting it in temporal contiguity with a reinforcer until conditioning had become strong. Then, a second stimulus (S2) was introduced at the same time as S1 and the reinforcer now followed both stimuli in the same temporal relation as before. For example, a tone would first precede food until salivation had been conditioned. Then, a light would be introduced and the tone and a light together would both precede food. Note that in this procedure, S2 stands in the same temporal relation to the reinforcer as S1, a temporal relation known to permit conditioning. If temporal contiguity were all that was required for conditioning, then S2 should also acquire control of the reinforcer-elicited response. However, tests showed that when S2 was presented by itself conditioning had not occurred. In terms of the previous example, the tone evoked salivation but the light did not. Prior conditioning to the tone blocked conditioning to the light even though the light had occurred in temporal contiguity with the food. This basic result was first clearly identified with the classical procedure but was soon replicated with the operant procedure. Various control procedures eliminated alternative interpretations of the blocking effect. The conclusion from this work was that a second factor in addition to temporal contiguity was required for selection by reinforcement. That factor is typically called discrepancy. For conditioning to occur, a reinforcer must not only appear close in time to the stimulus in Pavlov's procedure or to the response in Thorndike's procedure, but the reinforcer must also evoke a change in behavior. Returning to the example, salivation already occurred when the light was first introduced because the accompanying tone had already been paired with food. As a result, the introduction of food after the light did not produce a change in behavior that was large enough to produce selection by reinforcement. The discrepancy requirement may be stated as follows:
\[ \Delta R = \alpha f(R_{max} - R_{current}), \]

where \( \Delta R \) is a change in the strength of conditioning between the environmental and behavioral events, \( \alpha \) is the proportion of the total possible change that can occur as a result of one contiguous sequence of events, \( R_{max} \) is the maximum strength of the response to the reinforcing stimulus (e.g., food), and \( R_{current} \) is the present strength of the reinforcer-elicited response to the prevailing stimuli (e.g., the sight of the bar). In short, a change in behavior is some function \( f \) of the reinforcer-related discrepancy that accompanies contiguous events. Learning theorists differ somewhat among themselves as to how best to characterize the nature of the discrepancy, but almost all agree that both discrepancy and temporal contiguity are required for conditioning. The first formal statement of the discrepancy requirement was proposed by Robert Rescorla and Allan Wagner.

**Biological Analysis of Reinforcement**

Although additional research on reinforcement is needed at the behavioral level of analysis, enough is now known to warrant an effort to identify its biological mechanisms. The motivation for identifying the biological mechanisms of reinforcement is not only to understand the reinforcement process more completely but also to promote acceptance of reinforcement as the central process in the origins of complex human behavior: The account of evolution through natural selection was not generally accepted until its biological mechanisms (genetics) had been discovered some 70 years after Darwin’s initial proposal. If the parallel holds, the acceptance of selectionist accounts of individual behavior awaits the discovery of its biological mechanisms.

Events that initially function as reinforcers, such as food-evoked salivation, do so primarily as the result of natural selection. Such events are called *unconditioned reinforcers* because their reinforcing ability is not dependent on experience. Unconditioned reinforcers stimulate receptors that ultimately activate nerve cells (neurons) in the midbrain that diffusely liberate the neuromodulator dopamine. Axons from these midbrain neurons project widely to frontal cortex and other regions that are involved in the emission of behavior. Thus, the release of dopamine is in a position to affect the strengths of connections (that is, *synaptic efficacies*) between many neurons. Moreover, dopamine is known to facilitate long-lasting increases in synaptic efficacies between simultaneously active pre- and postsynaptic neurons. This facilitation is known as *long-term potentiation*, or LTP. LTP is believed to provide the neural basis of conditioning and was first identified by Timothy Bliss and Terje Lomo.

An experiment by Wolfram Schultz and his colleagues illustrates the role of dopamine in conditioning. A restrained monkey was presented with a light followed closely by a squirt of orange juice into its mouth while monitoring the activity of dopamine-liberating neurons in the midbrain. At first, these dopamine neurons were activated only by the juice, but after a number of light-juice sequences the light began to activate them as well. Once the light activated the dopamine neurons, the juice no longer did. The failure of juice to activate the dopamine-releasing neurons once the light had acquired this ability is the neural basis of blocking. The acquired ability of light to activate dopamine neurons also indicates that the light could now serve as a reinforcing stimulus. Stimuli that acquire the ability to serve as reinforcers after being paired with other reinforcers are known as *conditioned reinforcers*. The cumulative effect of experience establishes many stimuli as conditioned reinforcers. In this way, learning in experienced organisms becomes increasingly independent of unconditioned reinforcers.

**Reinforcement and the Emergence of Complex Behavior**

The acceptance of natural selection as the primary insight into evolution was also dependent on the development of quantitative methods to trace the cumulative effects of natural selection. Using the computational methods of population genetics, natural selection was shown to be competent to produce the complexity and diversity of species found in nature. Convincing demonstrations that selection by reinforcement provides similarly powerful insights into the origins of complex individual behavior require analogous methods. Among the more promising computational methods are *artificial neural networks*. Artificial neural networks are interconnected sets of units that simulate the interconnected
neurons that make up the nervous system. A goal of neural network research is to demonstrate that when the inputs to the network are activated in sequences that mimic the experience of an organism, the outputs of the network simulate the behavior observed in experienced organisms. The strengths of connections between units in artificial neural networks are modified by computational procedures that simulate the action of reinforcement and other experimentally identified processes on synaptic efficacies. Neural networks of interest to experimental science are those that are informed and constrained by findings from behavioral and biological research. Other fields that exploit neural networks, such as artificial intelligence, need not honor these constraints but are primarily concerned with whether the outputs of the network are effective for the task at hand. Exploring the implications of reinforcement by means of neural networks is in a relatively early stage of development but already shows promise with such complex behavior as language acquisition, concept formation, and memory. The work of James McClelland and David Rumelhart has been pioneering in this effort.

John W. Donahoe

See also Behaviorism; Decision Making and Reward, Computational Perspectives; Discrimination Learning, Training Methods; Distributed Cognition; Human Classification Learning; Memory, Neural Basis; Natural Action Selection, Modeling

Further Readings


The development of relationships with significant others is one of the most important tasks that an individual encounters in his or her lifetime. Relationships, according to Robert Hinde, are ongoing patterns of interaction between two individuals who acknowledge some connection with each other. In the case of children and adolescents, the social partners with whom interaction is most frequently experienced include parents, peers, and teachers. From Hinde’s perspective, individuals bring to social exchanges reasonably stable social orientations (temperament; personality) that dispose them to be more or less sociable and a repertoire of social skills for understanding the thoughts, emotions, and intentions of others and for interpersonal problem solving. Over the short term, a child’s or adolescent’s interactions with others will vary in form and function in response to fluctuations in the parameters of the social situation, such as the parent’s or peer’s characteristics, overtures, and responses. Often, social interactions are embedded in longer term relationships and thus are influenced by past and anticipated future interactions. For example, the nature of any given relationship is defined partly by the characteristics of its members and by its constituent interactions. Over the long term, the kind of relationship that any two individuals form with one another depends largely on the history of their interactions and relationships, not only with each other but also
with other members of their personal social community. Consequently, the first dyadic relationships that children experience are embedded within a group—the family. Significantly, families help define the type and range of relationships and interactions that are likely or permissible.

Many theories of human development (e.g., Sigmund Freud, Erik Erikson) suggest that relationships with others are important to healthy social and emotional development. This entry reviews the development of relationships with caregivers or parents, peers, friends, and romantic partners. The central argument presented herein is that the earliest relationships children form with their primary caregivers help shape the formation of internalized, mental representations of relationships, which, in turn, subsequently affect the development of other significant relationships.

**Attachment Relationships**

John Bowlby proposed that the attachment relationship between the child and his or her primary caregiver (most often, the mother) derives from a biologically rooted behavioral system that is marked by the infant's natural proximity seeking to caregivers for safety, security, and support. The attachment system regulates both physical and psychological safety in the context of close relationships. Perceived danger, stress, and threats to the accessibility of attachment figures activate attachment responses. When children with secure attachments are threatened, they tend to seek out those with whom they have formed attachments, and in this way, these figures serve as “safe havens.” In novel environments, attachment figures also serve as “secure bases” from which children explore their environment. Herein we briefly review important concepts about attachment and internal working models.

Infant attachment to caregivers is typically assessed through a laboratory paradigm developed by Mary Ainsworth and colleagues. Ainsworth’s *strange situation* comprises several episodes during which caregivers and strangers enter and leave an unfamiliar room within which the child is present. The task is designed to mimic how familiar and unfamiliar adults flow in and out of a child’s daily life. The quality of the attachment relationship is assessed by observing how the child explores the unfamiliar environment when the caregiver is present and how the child reacts to the departure and subsequent return of the caregiver. Questionnaires, interviews, and other observational paradigms have been developed to assess attachment in alternative settings and with older children and adults.

Secure attachments result from sensitive and responsive caregiving. The sensitive and responsive parent interprets signals correctly and responds effectively and appropriately to the child’s behaviors and needs. Sensitive and responsive parents do not direct anger or hostility to their young children, even when they are feeling irritated or annoyed. In the strange situation, infants who explore the environment freely, engage with strangers while the caregiver is present, and seek proximity to caregivers when under stress are classified as *securely attached*. When the caregiver leaves, securely attached infants are visibly upset, but on reunion, they are relieved to see the caregiver and easily soothed. Secure infants become children who express their emotions to others and actively seek help when they are unable to help themselves. These behaviors help them learn to regulate their emotions, adapt to new challenges, and develop healthy relationships with others during their lifetime.

When parents are insensitive and unresponsive, their infants develop insecure attachments to them. There are three types of insecure attachments: *anxious-avoidant*, *anxious-ambivalent*, and *disorganized*. In the strange situation, children who do not seek caregivers in times of stress and/or ignore caregivers after separation are classified as anxious-avoidant. Anxious-avoidant children often have caregivers who ignore or reject them in times of need. These children show limited affective engagement with caregivers, learning to inhibit their negative emotions and avoid emotional interactions. Anxious-avoidant children have difficulty controlling their anger in social company and thus have difficulty developing positive peer relationships and friendships.

Children who are unusually clingy with caregivers in the strange situation and need more reassurance than other infants, even in only mildly stressing situations, are typically classified as anxious-ambivalent. These children have more difficulty separating from parents, and during reunion they are more difficult to comfort. Anxious-ambivalent children have
caregivers who are inconsistent in their availability and sensitivity. Thus, these children display vigilance for caregiver actions and show inflated distress in order to elicit caregiver attention. In social interactions, these children are easily frustrated, impulsive, and overly anxious. They are less likely to explore in novel situations and have heightened personal fears. With peers, some of these children may act aggressively, whereas others act more passively and are prone to social reticence and withdrawal.

Last, children who develop disorganized attachment relationships are likely to have parents who are emotionally or physically abusive. These children show no clear pattern of behavior in the strange situation. They are at greater risk than other attachment groups to be aggressive and are more likely to develop oppositional defiant disorder, a persistent pattern of uncooperative, hostile behavior that interferes with a child’s basic functioning.

Note that the cross-cultural universality of attachment theory has been questioned. Critics argue that attachment theory emphasizes autonomy, independence, and individuation as defining competence, all of which are rooted in Western ideals. They also emphasize that caregiver sensitivity may be culturally defined and thus differ among societies. Consequently, traditional measures of attachment, such as the strange situation, may not be relevant in all cultures. For instance, in Eastern cultures, such as Japan, dependence and accommodation are encouraged in children. Furthermore, babies in those cultures generally experience less separation from their caregivers and subsequently may be more stressed by the strange situation than American babies. These cultural differences may explain, in part, why Japanese babies are more likely to be classified as insecure-ambivalent than are babies from the United States. However, with the acknowledgment that sensitive parenting, the secure base, and competence may differ in expressed form across cultures, attachment theory is still useful in understanding the power of parent-child relationships in later significant relationships.

Internal Working Models

Central to attachment theory is the proposition that early attachment relationships provide a basis for representations of self and others in social relationships beyond the parent-child dyad. These \textit{internal working models} (IWMs) are hypothesized to become so deeply ingrained that they influence feelings, thoughts, and behaviors with significant others at both the conscious and unconscious level. For example, if the caregiver is consistently responsive to the child’s needs, she or he will feel confident, secure, and self-assured when introduced to novel settings. In contrast, if the caregiver is unresponsive to or rejecting of the child, she or he may develop an IWM about the self as being unworthy of care; interpersonal relationships are expected to be rejecting or neglectful, and the social world is viewed as hostile and unwelcoming. Thus, early relationships affect a child’s “felt security,” a significant developmental phenomenon that provides the child with sufficient emotional and cognitive sustenance to allow for the active exploration of the social environment.

In accord with their internalized expectations about the social world, children act in ways that confirm these beliefs. For example, if a child has negative expectations of peers, he or she is likely to think that an ambiguously intended harmful event was intentionally caused. The child is also more likely to choose to react in aggressive ways toward the perceived provocateur, increasing the likelihood that the peer will respond in a hostile manner in return. Like a self-fulfilling prophecy, IWMs cause individuals to behave in ways in which others fulfill their expectations, strengthening and reinforcing their original schemas. Consequently, IWMs become more difficult to modify with increasing age.

Friendships

Consistent with the assumption that IWMs are carried forward from relationship to relationship, significant associations exist between security of attachment in the parent-child relationships and the quality of children’s close dyadic friendships. Friendships typically comprise the first significant nonfamilial relationship that children develop with others. Friendships may be defined as reciprocal, egalitarian relationships in which both partners acknowledge the relationship and treat each other as equals. Friendships are typically characterized by companionship, a shared history, and mutual affection.

Children with secure attachments to parents have more friends and their friendships are of better quality than those of insecurely attached children. Indeed, interactions between friendship dyads comprising
two securely attached members are more positive, fair, intimate, and responsive than interactions within dyads comprising only one securely attached member. Moreover, securely attached adolescents are viewed by their best friends as being more altruistic and more conciliatory after conflict; also, they are more satisfied with their friendships than the friends of anxious-avoidant or anxious-ambivalent adolescents. While there generally are associations between parent-child attachments and later peer and friend relationships, a child’s attachment relationship to parents is not absolutely deterministic of their later relationships with friends. There are children who are insecurely attached to parents and yet form high-quality friendships. In this way, a good friendship may compensate for the child’s insecure attachment to parents.

Wyndol Furman has interviewed young adolescents to explore their IWMs of friendship. Adolescents classified as having secure working models of friendship, or who recounted their relationships in a coherent way and reported that their friendships were influential and valuable, reported more warmth and support in their relationships. Adolescents who were categorized as having dismissing IWMs of friendship, or who had little interest in caregiving and support seeking from friends, also reported that their friendships were unsupportive and lacked warmth; individuals with preoccupied IWMs of friendship who described overconcern for their friends’ problems and were vague, angry, or passive in their descriptions of their friendships were more likely than the other groups to be in relationships that had power imbalances.

Finally, the quality of early parent-child attachment relationships predicts the quality of relationships with friends; this association is stronger for older children and adolescents than for younger children who may still rely on parents as their primary attachment figures. As Ainsworth suggested, attachment working models seem to be more influential for friendship development and maintenance as children enter adolescence, when intimacy and social support become more central features of friendship.

**Peer Relationships**

Given an IWM that the parent is available and responsive, the young child feels confident, secure, and self-assured when introduced to novel settings. Thus, felt security provides the child with sufficient emotional and cognitive sustenance to allow the active exploration of the social environment. Exploration results in active and interactive play, which, in turn, leads to the development of social competence and interpersonal problem-solving skills. From this perspective, there is a clear association between security of attachment in infancy and the quality of children’s social skills and competencies. Indeed, because they demonstrate socially skilled behavior, securely attached children are generally accepted and liked by their peers.

Alternatively, the development of an insecure infant-parent attachment relationship appears to result in the child’s developing an IWM that interpersonal relationships are rejecting or neglectful. In turn, the social world is perceived as a battleground that must either be attacked or escaped from. For the insecure and angry child, opportunities for peer play and interaction are nullified by displays of hostility and aggression in the peer group. Such behavior, in turn, results in the child’s forced (by the peer group) lack of opportunities to benefit from the communication, negotiation, and perspective-taking experiences that will typically lead to the development of a normal and adaptive childhood. For the insecure and wary or anxious child, opportunities for peer play and interaction are nullified by the child herself or himself. Consequently, social and emotional fearfulness prevail to the point at which the benefits of peer interaction are practically impossible to obtain. Thus, because they demonstrate socially unskilled aggressive or fearful behavior, insecurely attached children are often rejected by the peer group.

**Romantic Relationships**

Like friendship, romantic relationships are voluntary, reciprocal, and egalitarian associations that provide partners with companionship, intimacy, and support. Unique to romantic relationships, partners are attracted to one another, share feelings of love, and engage in sexual behaviors. Romantic relationships also differ from friendships in that they become more obligatory and exclusive over time, especially if the relationship is publically formalized.

Empirical links between security of attachment in infancy and early childhood and subsequent romantic relationships are beginning to emerge. Preliminary evidence suggests that early secure
attachment predicts more positive feelings, felt security, and support, and less negative behavior in romantic relationships over 20 years later. These associations are often indirect, being explained, in part, by social competence and peer acceptance during childhood and secure friendships in adolescence. Additional research has indicated that securely attached adolescents interact positively with romantic partners, even if they were not securely attached to mothers as infants.

In summary, researchers have revealed stability between individuals’ attachment representations with caregivers in infancy and childhood, their friendships in childhood and adolescence, and their romantic relationships in adulthood. Empirical examinations of these representations suggest that individuals have both a general working model of others as well as domain-specific representations of relationships (e.g., parent-child relationships, friendships, and romantic relationships).

Kenneth H. Rubin and Kristina L. McDonald

See also Love; Personal Identity, Development of; Self, Development of; Social Cognition

Further Readings


**RELIGION AND PSYCHIATRY**

Religion is an organized system of beliefs, practices, and rituals designed to facilitate closeness to the sacred or transcendent—whether that be God or a higher power (in Western traditions) or ultimate truth or reality (in Eastern traditions). Religion includes specific beliefs and personal commitment to those beliefs, which reflects their overall religiousness and religious motivation (the degree to which religious beliefs and goals are the person's ultimate concern in life). Another essential aspect of religion is its emphasis on one’s relationship with and responsibility to others living together in a community that shares common beliefs, rituals, and practices. Besides attending religious services and other forms of involvement in religious community activities, the religious person may also be involved in private religious activities such as prayer, the reading of sacred scriptures, and forms of worship and ritual that are performed when alone.

Psychiatry is the specialty of physicians who receive special training to treat the many forms of mental illness that disrupt a person’s sense of peacefulness, hope, and meaning in life. That disruption is often so severe that the person is no longer able to function in healthy ways in their social interactions, work, or recreational activities. Although psychiatry is mainly concerned with the negative or dysfunctional aspects of mental health (depression, mania, anxiety, psychosis, personality disorders, substance abuse, etc.), it also seeks to enhance the positive side of mental, emotional, and behavioral health. Positive mental health has to do with happiness and well-being, being satisfied with life, and having joy, peace, hope, optimism, meaning, and purpose as one pursues life’s goals. It also involves having satisfying long-term social relationships and engaging in work that is productive and meaningful. Psychiatry is also concerned with mental states that fall in between positive mental health and mental disorder. This includes helping people deal with a life that, while not dysfunctional, has become unsatisfying, boring, meaningless, or hopeless. This entry focuses on the relationship of religion to mental illness, mental health, and well-being.

**Historical Background**

For the last 100 years or so, religion has been viewed within psychiatry as related to neurosis
and unhealthy functioning. Indeed, Sigmund Freud described religion as an “obsessional neurosis” and psychologist Albert Ellis suggested that the less religious that people are the more emotionally healthy they will be. Other psychiatrists have emphasized that religion is incompatible with mental health, adversely affects self-esteem, self-actualization, and mastery and disrupts healthy sexual functioning. Such opinions, however, are not derived from systematic research but rather from personal experiences and clinical exposure to patients with mental illness who often express their religion in pathological ways.

**Recent Research**

Within the past two decades, however, systematic research has begun to examine the relationship between religious involvement and mental health in surveys of community populations and persons with physical rather than mental illness. It has been discovered that religion is often used to cope with stresses involving loss of health, loss of loved ones, or other traumatic losses. Harold Koenig and his colleagues have summarized this research in a number of books and articles. These reviews have uncovered over 1,000 studies that have examined relationships between religion and both negative mental health (depression, anxiety, etc.) and positive mental health (happiness, well-being, hope, meaning and purpose, etc.). Nearly two thirds of these studies have found that the religious person on average experiences better mental health, fewer negative emotions, greater social support and is less likely to be engaged in substance abuse. Most of this research has come from the United States, but studies reporting similar findings have been conducted around the world, including Canada, South and Central America, the United Kingdom, Europe, the Middle East, and other continents. While most of these studies are epidemiological in design (cross-sectional or prospective observational studies), a number of randomized clinical trials have been conducted in patients with depression and anxiety disorders, finding that religious interventions often result in faster improvement compared to traditional secular psychotherapy or no treatment. These studies include interventions from a variety of faith perspectives, including Christian, Muslim, and Buddhist approaches.

Besides studies showing lower rates of depression, faster recovery from depression, and faster response to religious therapies, an even larger research base shows relationships between religious involvement and positive emotions. Of the more than 350 studies that have now examined relationships between well-being or happiness and religious involvement, over three quarters report that the religious person experiences significantly more of these positive emotions than the less religious or nonreligious person. With regard to substance abuse, the results are similar. In the over 375 studies that have now been done, over 80% indicate that the religious person is less likely to drink, use drugs, or smoke cigarettes. These studies have often been done in young persons whose entire lives are ahead of them. Better mental health and healthier lifestyles translates into better physical health as well, which itself influences mental health.

**Explanations**

Why is this so? How might religion enhance mental health? First, religion can provide a positive, optimistic worldview that gives life meaning and purpose and provides hope. Second, religion often provides rules and regulations to help guide behavior. Third, religion may enhance social interactions by its emphasis on forgiveness, thankfulness, gratitude, generosity, and other attitudes and activities that foster healthy interpersonal relationships. As a result, religion not only helps people to cope but may also reduce the amount of stress that they must cope with.

**Not Always Positive**

Despite the many contributions that religion makes to mental health, this does not mean that the religious person will always be healthier and happier than the less religious or secular person. Religion can be at times associated with excessive guilt, anxious ruminations, obsessions, prejudice, and can therefore lead to a restricted and limited life, rather than a fuller and freer one. People with psychological problems often turn to religion for comfort, and while religion helps them cope, it may not always eliminate the personality disturbances or the inherited genetic vulnerabilities to mental illness that continue to persist, although moderated to some extent by religion. Religion, then, can be used in neurotic or mentally unhealthy ways to justify actions, judge others, or become the object of ruminations and obsessions. Pastoral counselors—that is, those who are counselors at the master’s or PhD level who also have religious education (seminary, divinity
school, or clinical pastoral education)—are trained to address these issues. Pastoral counselors are different from community clergy, who may have little or no training in dealing with mental health issues.

**Conclusion**

In the balance, though, religion is an important resource for mental health, a resource that most mental health professionals—including psychiatrists—have ignored for a long time. In this day and age with research accumulating, that is no longer possible.

*Harold G. Koenig*

**See also** Behavioral Therapy; Emotion and Psychopathology; Happiness

**Further Readings**


---

**Representational Theory of Mind**

According to the representational theory of mind (RTM), mental representations are often involved when we have a mental state or engage in a mental process. Mental representations are symbols that exist in the mind. Being symbols, they have semantic properties; that is, they have meaning or content and so are about particular things or states of affairs. This entry provides a description of the most prominent historical and contemporary versions of RTM along with a description of the key debates surrounding RTM.

**Historical Versions of RTM**

RTM was prominent in the modern period of philosophy of the 17th and 18th centuries, being associated with such philosophers as René Descartes, John Locke, David Hume, and Immanuel Kant. Advocates of the theory in this period tended to advance RTM as a theory about such familiar mental states as beliefs, desires, intentions, and so on (mental states generally known as propositional attitudes), and about mental processes of thinking involving such states. They were also prone to regard mental representations as being images that are introspectable, private, and immaterial.

**The Return of RTM**

With the rise of behaviorist views of the mind in both philosophy and psychology, traditional versions of RTM fell out of favor in the early decades of the 20th century, particularly in the English speaking world. However, as the limitations of behaviorism became apparent a new version of RTM was developed that came to dominate the newly emerging field of cognitive science in the 1960s and 1970s. Such a view has been given its clearest articulation and most thoroughgoing defense by the American philosopher Jerry Fodor. For Fodor, the representations involved in having a propositional attitude such as a belief or a desire are language-like rather than imagistic so that, for example, believing that dogs bark involves having a sentence in one’s mind that means *dogs bark*. Fodor labels the language that the mind employs the *language of thought* (LOT). LOT is a nonnatural language that will be shared by all members of the human species regardless of what language they speak. Being a language, LOT has a finite number of basic symbols and a finite number of syntactic rules for combining those symbols to create larger complex structures such as sentences. The meaning of a sentence of LOT is a product of the meaning of its component words and its syntactic structure.

Fodor is a physicalist in the respect that he thinks that the mind is ultimately a physical thing whose mental properties are determined by its physical properties. Consequently, he thinks that the symbols of LOT are physically embodied in the brain. However, any given LOT sentence is multiply realizable in the sense that its instances can take a variety of different physical forms in the brain of distinct individuals. For Fodor, whether or not an instance of a particular sentence of LOT in one’s mind
expresses a belief, a desire, or whatever, depends on how it is processed by the mental mechanisms that have access to it. Sentences that express beliefs are processed in the distinctive way that is characteristic of the belief relation and so on for all the other types of propositional attitude relations. This idea is often figuratively expressed by saying that when one has a particular belief, one has a relevant sentence of LOT in one’s belief box; that when one has a particular desire, one has the relevant sentence of LOT in one’s desire box; and so on.

Fodor supplements his theory about propositional attitudes with a theory about mental processes. According to this theory, mental processing involves the manipulation of symbols of LOT by means of computation. Hence, the mind is a computer. A computer, on Fodor’s conception, is a symbol manipulating system that takes syntactically structured symbols as input and generates syntactically structured symbols as output by means of the application of symbol manipulating rules. Although the symbols have semantic properties so that a computer’s activity can be characterized in semantic terms (as processing information or solving problems), the computer will have no access to those semantic properties.

Fodor argues that his version of RTM, unlike its competitors, explains several salient facts about the mind. Prominent among these is that thought is systematic in that anyone capable of believing that object a stands in relation R to object b (for example, that John loves Jill) is also capable of thinking that b stands in relation R to a (for example that Jill loves John). Fodor has also argued that his version of RTM has scientific support in that it underlies most mainstream work in cognitive science. This claim was perhaps true when first made by Fodor in the 1970s and reflects the fact that the theory can be extended (as it was by many cognitive scientists) to apply to mental states and processes not recognized by commonsense psychology, including those that reside at the unconscious or subpersonal level. However, the 1980s witnessed the rise of an alternative connectionist approach that is widely adopted within contemporary cognitive science.

The Connectionist Challenge

According to connectionism, mental processing is supported by the activity of neural networks consisting of simple units that are connected to one another so that the activation of one unit can be communicated to other units in the network, so pushing them toward a state of activation. In this way, a typical connectionist network serves to transform patterns of activation at an input layer of units into patterns of activation at an output layer. Because the patterns of activation represent things or states of affairs, the network, just like an orthodox computer, processes information or solves problems. Hence, connectionism can be viewed as a version of RTM, although many connectionists would resist such a characterization as they are keen to emphasize their opposition to traditional versions of RTM such as Fodor’s. The representations processed by a connectionist network do not typically have syntactic properties. Moreover, because the network’s “knowledge” is stored en masse over the connections between its constituent units, it does not store information by means of syntactically structured symbols. For his part, Fodor has objected to this new version of RTM by arguing that it cannot explain the systematicity of thought.

The Chinese Room Argument

One of the most widely discussed philosophical objections to Fodor’s version of RTM is the Chinese room argument devised by John Searle. Searle, who has no grasp of Chinese, imagines himself trapped in a room containing batches of sheets of paper with Chinese symbols written on them. The room also contains a book written in English instructing Searle how to correlate symbols of Chinese with symbols of Chinese. Further sheets with Chinese symbols written on them are posted into the room. Searle responds to this input by executing the English instructions that involve considering the syntactic properties of the input symbols and correlating them with symbols written on the batch of sheets. These symbols are then copied onto blank sheets that are in turn posted out of the room. The input symbols are actually questions written in Chinese and the output symbols are sensible answers to those questions so that Searle’s symbol processing behavior mimics that of a competent speaker of Chinese. Searle’s point is that he does exactly what a computer does, yet he doesn’t understand Chinese. Hence, he concludes, no computer, however it is programmed, is capable of understanding Chinese or any other language, a conclusion that he generalizes to all cognitive capacities.

Searle’s argument has been widely discussed but no general consensus as to its power has emerged. A common response made on behalf of RTM is the
so-called systems reply, a version of which can be described in the following terms. Advocates of RTM are not committed to the claim that computation is sufficient for cognition, only that computation plays a fundamental role in our mental lives. If we should accept that Searle in the room does not understand Chinese, this is because he simulates only a limited element of the behavior of a Chinese speaker. In particular, he does not respond to nonsymbolic input. For example, if a ferocious dog were let into the room Searle wouldn’t produce the Chinese equivalent of help or get me out of here. Moreover, Searle does not respond to Chinese symbols by engaging in relevant nonsymbolic behavior. For example, if a note were posted into the room saying, “There is a bomb under your chair, and to defuse it you need to invert your chair” in Chinese, he would not respond by inverting the chair. Now suppose we built a robot containing a powerful computer. The computer is fed information from a video camera attached to the robot and issues instructions that cause the robot to move around its environment and manipulate objects. In short, the computer is hooked up to the robot’s perceptual and motor systems. Also suppose that the computer is programmed in such a way that it responds to Chinese symbols in a way that is coherently related to their meaning and that this sometimes involves engaging in nonsymbolic behavior. Moreover, suppose that it sometimes produces appropriate symbols of Chinese in response to nonsymbolic impingements. Then perhaps the robot as a whole (as opposed to any of its internal subsystems) would understand Chinese. And its ability to understand Chinese would be a product of its computational activity so that that activity would have to be appealed to in order to explain how the robot understands.

Conclusion

In sum, RTM constitutes an enduring theory as to the nature of mind and cognition. Its contemporary versions, particularly as developed and defended by Fodor, are currently both prominent and popular. However, they are subject to potentially powerful philosophical and empirical challenges.

Mark John Cain

Further Readings


Representations, Development of

The main focus of this entry is the development of conceptual representations, in particular the representation of objects, spatial relations, and events. A secondary focus is the differentiation of conceptual from perceptual representations. By conceptual representation is meant the construal or meaningful interpretation of perceptual and linguistic information. A common assumption in the field is that such interpretations are explicit (i.e., have the potential to be brought to conscious awareness), whereas it is known that many aspects of perceptual representations are implicit (i.e., cannot be brought to awareness). In the mature organism, perception is suffused with conceptual interpretation, and it is difficult to disentangle the two, but the two kinds of representation follow somewhat different developmental courses. Infants come with few, if any, interpretations of the world but rapidly form perceptual categories of what objects look, sound, and feel like and the kinds of movements they make or are made with them. In some cases, quite detailed perceptual categories appear by 3 months of age. It is not known exactly when conceptual interpretation begins, but it is in evidence at least by 6 months and possibly considerably earlier. In contrast to perceptual categories of objects, early object concepts tend to be global or general in nature. They slowly differentiate into finer grained concepts, but the initial concepts form the bedrock of the conceptual system and remain throughout life.

Early Object and Event Concepts

The traditional view of concept formation, that of Jean Piaget, was that the first 1½ years of life are
a period of exclusively sensorimotor development, in which infants learn to recognize objects and the daily events of life and to respond appropriately to them. Conceptualization was said to develop slowly from these perceptual and motor routines, eventually allowing infants to think about objects or events in their absence, to recall the past, to solve problems mentally, and to imagine the future. However, experimental findings in the last two decades have shown that conceptual activity begins much earlier in life, perhaps as early as a few months of age. Infants do learn to recognize and respond appropriately to objects and events in the first 2 years, often in quite a detailed way. For example, they can perceptually differentiate dogs from cats as early as 3 months. However, they also begin to conceptually interpret objects and events quite early in this period, albeit in a more global, less detailed fashion.

A matter of controversy in the field is what, if any, innate proclivities may be required to begin interpreting perceptual input. One proposal is that certain core knowledge is innate, such as that objects move on continuous paths and are solid in the sense that two objects cannot occupy the same space. A somewhat different proposal is that core concepts need not be built in but various aspects of perception, particularly motion through space, are preferentially attended from birth and are redescribed in simplified format to create the first conceptual representations. Still another proposal, often associated with connectionist learning models, is that conceptual knowledge can be derived through perception itself without any innate biases or redescription into another format.

By around 6 months of age, conceptualization is shown by infants beginning to recall absent objects and events. Evidence for conceptual representation of events is shown by deferred imitation, a form of nonverbal recall in which observed events are reproduced after a delay. Recall of a past event requires that a conceptualization of it be brought to awareness. Around the same age, evidence for global object concepts such as animal and vehicle is shown by familiarization-dishabituation studies, in which infants are given several little models of animals to handle and then are given a new animal or a vehicle. By 7 months, infants show global object concepts by dishabituating to (increasing their interactions with) any vehicle after interacting with animals (and vice versa). Visual similarity of shape aids global conceptualization but is not essential, as shown by 9-month-olds differentiating little models of birds and airplanes, all of which are quite similar in appearance. Around this age infants also broadly generalize from one instance of a class to another, as shown by their being willing to substitute new members from the same conceptual class when imitating. For example, after seeing an event in which a little model of a dog is given a drink, infants as young as 9 months will haphazardly choose a little model of another dog or any other animal to imitate this event but will not choose a vehicle or other artifact.

Because of the prevalence of Piagetian theory, experimental study of infant conceptual representations began relatively recently. However, research findings such as those just described indicate that the course of the development of object concepts over the first 2 years tends to begin at a global or superordinate-like level, such as animal, vehicle, furniture, and plant and then gradually differentiates to more detailed concepts, such as dog, car, chair, and flower. The onset of language understanding near the end of the first year contributes to this differentiation, because the language children hear is more differentiated than the sketchy concepts they first bring to the language-learning task and thus emphasizes details that previously may be unattended. The initial global concepts, however, organize the further learning that conceptually differentiates one animal or artifact from another, thus leading to a hierarchical system of object concepts. This organization lasts throughout life, barring brain damage. The way the conceptual system breaks down in semantic dementia testifies to the foundational nature of global concepts such as animal; detailed information is lost first and global information is the longest lasting.

The first object and event concepts appear to be heavily influenced by spatial information, especially motion through space. From birth, infants are attracted to motion and in the early months are more apt to notice that something moves and even the kind of path that it takes than what it looks like. Hence, differences in motion, as well as contingent interactions between objects (as in peekaboo games or goal-directed actions), are likely bases for the initial concepts of animal and nonanimal. One view is that a first conceptualization of animal is a thing that starts motion by itself and interacts with other objects from a distance, whereas a nonanimal either doesn’t move at all or, if it does, doesn’t start motion by itself and doesn’t interact with other objects from a distance. Again, simple notions such
as these remain with us throughout life, even when sometimes contradicted by more detailed biological knowledge.

**Early Relational Concepts**

It is more difficult to differentiate early concepts of spatial relations from perception of them. However, extensive research on relations such as containment, support, and occlusion suggests that at least by 3 to 4 months of age infants are beginning to conceptually interpret these relations. For example, 3-month-olds act as if they expect that when an object goes behind a screen or other occluder it will be hidden; hence, they dishabituate (look longer) if there is a window in the occluder and the object comes into view as it passes the window. Similarly, infants this age look longer at a wide object moving behind a thin occluder when it sticks out of either side than if it is completely hidden. It does not seem plausible that such expectations could be taught by perception alone but instead seem to require some interpretation of what is perceived. The conceptualization of spatial relations, like that of objects, begins in a general or global fashion and gradually becomes more detailed. For example, a concept of containment appears to begin without any quantitative understanding of the relationship between height or width of a container and what it can contain. Such variables are gradually learned over the course of the first year.

Spatial relational concepts are somewhat more subject to linguistic restructuring than are object concepts. Language helps differentiate global object concepts (e.g., children hear the words *dog* and *cat* more often than *animal*, and *car* and *truck* more often than *vehicle*) but rarely restructures object domains. In contrast, spatial relational concepts may be at least partially restructured by language. For example, Korean distinguishes degree of fit (tight versus loose), whether the fit is of a containment or support relation; for example, the same word is used to describe a ring on a finger as a finger in a ring. Thus, the Korean language carves up the spatial domain in a somewhat different way than does English, and Korean-speaking children begin to differ in their spatial categorizations from English-speaking children by the end of the second year. In spite of the fact that both cultures have concepts of containment, support, and tight fit, these differences in the way they are accustomed to think about them remain in adulthood.

**The Developmental Course of Conceptual and Perceptual Representations**

In the sense of developing appropriate expectations and actions vis-à-vis objects, young children typically know more about objects and events than they can express, and this is only partly due to limited vocabulary. Although even young infants develop many expectations about how objects behave in the world and also learn how to interact with them successfully, this does not necessarily imply conceptual understanding of that behavior. This divergence between growth of perceptual knowledge and conceptualization is at least partly responsible for what are known as U-shaped developmental curves, in which infants are successful at some tasks while toddlers fail and older children are once again successful. For example, if infants are shown an object dropping to the floor and then a shelf is placed between the object and the floor and a screen placed in front so they cannot see the actual landing place, they expect the object to remain on the shelf rather than continue down to the floor. However, in similar situations 2-year-olds will search for the object on the floor rather than on the shelf. Prediction as a task requirement is a conceptual task, requiring activation of a conceptual representation of the physical world, which is different from perceptual expectations of how the objects behave in the world. Discrepancies between the two kinds of representation often continue into adulthood, indicating that these two forms of representation can exist throughout life without one always influencing the other.

_Jean M. Mandler_

**See also** Concepts, Development of; Event Memory, Development; Knowledge Acquisition in Development

**Further Readings**


**Representativeness Heuristic**

This entry provides a brief explanation and background for the representativeness heuristic, a cognitive process hypothesized to underlie people’s intuitive judgments of probability.

**Heuristics and Biases**

The representativeness heuristic is a theoretical construct that forms part of the influential heuristics and biases framework for explaining intuitive judgment in humans pioneered by Daniel Kahneman and Amos Tversky in the 1970s. From this perspective, because of their limited time, knowledge, and computational ability, in general people cannot make judgments according to the often complex normative rules of logic, probability theory, and statistics. Instead, they have to resort to simpler judgment heuristics that exploit natural assessments. Natural assessments benefit from naturally existing propensities of memory and perception that are conveniently available and easily assessed. These judgmental heuristics are often useful, but sometimes they lead to systematic and serious cognitive biases.

In regard to confidence (probability) judgment, the key notion has been variable substitution, according to which the complex judgment of probability is substituted with a simpler natural assessment. Assessment of probability in the sense implied by probability theory and statistics involves consideration of all possible outcomes and their frequencies of occurrence. In the face of this complexity it is proposed that people substitute probability with a subjective variable that is conveniently available and easier to assess. The representativeness heuristic suggests that the degree to which an instance or event is representative of a category is used as a proxy for the probability that the instance or event belongs to the category, as when you assess the probability that a person is a lawyer by assessing how similar he or she is to your stereotype for a typical lawyer.

Although representativeness is a useful guide to probability in many real-life circumstances, because it does not obey the rules of probability theory, use of the heuristic is claimed to produce a number of biases, or cognitive illusions, in probability judgment. Two classical demonstrations of such biases are base-rate neglect and the conjunction fallacy.

**Base-Rate Neglect**

If you are in Florida and encounter a male with a short haircut driving a Chevrolet Corvette, you may consider the probability that he is an astronaut (assuming that a clean-cut male driving a sports car close to Cape Canaveral fits your stereotype for an astronaut). In this case, probability theory implies that you should take into account the base rate of astronauts, which presumably is low also in Florida, and the modest reliability of hairstyle and car make as predictors of profession (per the celebrated Bayes’s theorem of probability theory). Because people use the representativeness heuristic when they assess the probability that the person is an astronaut, it is proposed that they only take into account the similarity between the person and their stereotype for an astronaut, while neglecting the base-rate of astronauts. Because of this base-rate neglect, they get too
captivated by the representativeness of the evidence. Likewise, even physicians tend to be too captivated by a positive result of a medical diagnosis test (e.g., for HIV), neglecting the often low prevalence (base-rate) of the disease in the population, thereby overestimating the probability that the patient actually has the disease.

The Conjunction Fallacy

Another classic judgment bias that is claimed to derive from use of the representativeness heuristic is the conjunction fallacy. Consider the following person description:

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and also participated in antinuclear demonstrations.

What is the probability of each of the following?

1. Linda is a bank teller.
2. Linda is a bank teller and is active in the feminist movement.

As noted by Tversky and Kahneman in 1983, most people assess the conjunction that Linda is a bank teller and a feminist (B&F) to be more likely than one of its constituents, that Linda is a bank teller (B). This violates a basic rule of logic and probability theory, which states that a conjunction can never be more probable than one of its constituents. More prosaically, it can never be more probable that a person is both a bank teller and a feminist than that the person is a bank teller per se, because the former set (B&F) is a subset of the latter (B), and any person who is a member of the B&F is necessarily also a member of B, but the reverse does not hold. The account proposed by the representativeness heuristic is that Linda is perceived as very unrepresentative of the category bank tellers and very representative of the category feminists, while the conjunction is perceived to be of intermediate representativeness. People will therefore assess a high probability that Linda is a feminist, a low probability that she is a bank teller. The probability of the conjunction falls in between these two constituent probabilities, leading to a violation of probability theory, in essence because similarity judgments do not obey the rules of probability theory.

It is clear that people’s probability judgments are often affected by perceived similarity relations and that people robustly produce these and a number of other cognitive biases, relative to the rules of probability theory. However, throughout the years, the heuristic has also been criticized, for instance, for being a too vague construct, and its exact role in producing these judgment biases is still an area of active research in psychology.

Peter Juslin

See also Availability Heuristic; Belief and Judgment; Debiasing; Similarity

Further Readings


**Resentment**

This entry defines the emotion of resentment, contrasts it with envy and r**essentiment**, links it with research on relative deprivation, and discusses its consequences. Resentment is an emotion we feel when we suffer a perceived wrong. It can be a powerful, motivating state, characterized by a blend of anger, bitterness, and indignation. The hallmark of resentment is that people feeling it believe that they have a justified moral complaint against another person or general state of affairs. They believe they have suffered undeservedly. Consequently, they feel resentment.

Resentment Contrasted With Other Emotions

It is useful to contrast resentment with envy. Envy involves a painful awareness of another person’s desired advantage and the blend of discontent, ill
Resentment will, and resentment that this awareness can produce. Thus, some sense of resentment or sense of injustice seems to be a common ingredient of envy. However, scholars emphasize that the resentment found in envy is highly subjective because it lacks social approval. Furthermore, it often results from a need to rationalize the ill will associated with the emotion. In its purest form, resentment follows a clearer-cut, seemingly objective, injustice and enjoys greater social approval. In some cases of objective injustice, resentment can appear as moral outrage or righteous indignation as was the case in the race riots of the 1960s and 1970s in the United States. Like envy, resentment can be fueled and exaggerated by rationalization. The wrong may even be imagined. In some cases, it may have originated from envy. But compared to envy, it is less likely to spring from a questionable starting premise.

Another important distinguishing feature of resentment is that, unlike envy, it need not arise from a social comparison. Much of the social science research on resentment focuses on people's reactions to disadvantage, but the range of situations that cause resentment is actually much more than disadvantage. We can resent being ignored when we are entitled to have a say in a group decision, for example. We can resent an insult or injury. We can resent tax rates or university parking policies. In other words, we can resent both unfair procedures as well as unfair distributions. Envy, however, is nonsensical without an explicit social comparison.

Finally, the action tendencies associated with resentment are more evident than with envy. When people feel resentment, because they perceive that they have been unfairly treated, they are liable to take action to remedy the wrong. Examples of open political violence have been explained by noting the resentment caused by group members being deprived of something to which they feel entitled and deserving. Envy can lead to actions, but because the emotion is socially repugnant and unsanctioned, these actions are more likely to be covert.

It is also useful to distinguish resentment from ressentiment, an emotion derived from Nietzschean ideas and further developed by another German philosopher, Max Scheler, in the early 20th century. Ressentiment refers to a state of mind resulting from chronic impotence and inferiority. It entails a devaluing of what one secretly craves but cannot obtain. It is like resentment in that it is a negative emotion often containing anger and frequently linked to deprivation. However, unlike resentment, it is passive rather than active. Ressentiment, generally, leads to self-debilitating inaction as a means of numbing the pain of inferiority. In contrast, resentment often leads to action to redress the perceived wrong.

Relative Deprivation

In social science research, resentment is closely linked with the broad topic of relative deprivation. People feel relatively deprived when another person (egoistic relative deprivation) or group (fraternalistic relative deprivation) enjoys a relative advantage, especially an advantage that prompts rising expectations for oneself or one's group. Resentment occurs when people feel entitled to and deserving of this advantage. An example of egoistic relative deprivation would be resentment because a fellow employee receives a promotion to which one also feels entitled. An example of fraternalistic relative deprivation would be resentment because a member of another racial group gains unfair admission into a professional program, thereby taking the place potentially enjoyed by a member of one's own group. Research on relative deprivation, with resentment often being the signature response, has a long and rich history in psychology, sociology, politics, and economics.

Consequences of Resentment

Recent studies link deservingness with resentment and then to subsequent schadenfreude, or pleasure derived from another person's suffering. For example, people find misfortunes suffered by high status individuals to be pleasing. This pleasure seems largely explained by the initial resentment felt because the high status is often perceived as undeserved.

Additional research links the resentment felt toward hypocrites and the special pleasure that arises when they are exposed for their hypocrisy. It may be that hypocrites, by their moralizing statements and “holier than thou” demeanor, amount to moral reproaches from the perspective of those around them. This creates a penetrating form of resentment because the moral core of the observers is threatened. By the same token, when hypocrites suffer exposure, the moral table is turned upside down. A self-threatening “upward comparison” is transformed into a pleasing, self-boosting “downward comparison.” Initial feelings of resentment in
observers seem to enliven the subsequent schadenfreude, even producing a sense of poetic justice.

Resentment is often intense and can lead to extreme, sometimes violent actions. Because people feeling resentment believe they have been unjustly treated and wronged, they can correspondingly feel justified in redressing the wrong. However, the possibility of biased, exaggerated, construals of the wrong mean that the redressing actions are themselves wrong. Spiraling retaliatory actions can then ensue. Scholars speculate that many intergroup conflicts mirror this pattern. Clearly, resentment is an important human emotion that deserves careful and sustained study.

Richard H. Smith and David Ryan Schurtz

See also Emotion, Cultural Perspectives; Envy; Intergroup Conflict; Jealousy

Further Readings


**Retrieval Practice (Testing) Effect**

The testing effect is a term used to describe the finding that taking a test on previously studied material leads to better long-term retention relative to restudying the material or not taking a test. Testing is often conceptualized as a neutral event in which the contents of memory are examined but left unchanged. However, the act of retrieving information from memory actually alters the retrieved memory by elaborating on the existing memory trace and/or creating additional retrieval routes. One consequence of these changes is that the probability of successful retrieval in the future is increased, making testing a potent mechanism for enhancing long-term retention. This entry provides a brief history of testing effect research followed by a discussion of the generalizability of the effect, potential theoretical explanations, and factors that increase its efficacy.

**History of Research on the Testing Effect**

The idea that retrieving information from memory can increase retention has a long history. Philosophers and other scholars have long recognized the mnemonic benefits of retrieval practice: Aristotle, Francis Galton, and William James, among others, all described how repeatedly recalling information from memory improves its retention. Some early studies in the 1900s confirmed their pronouncements, but only recently have researchers investigated this phenomenon systematically. The early studies were conducted with students in classroom settings and to demonstrate that testing improved retention of course material. In the following decades, research on the testing effect was sporadic. Most studies during this period were part of the verbal learning tradition and investigated the memorial consequences of retrieval in laboratory settings using discrete verbal materials, such as lists of individual words or word pairs. Since the start of the 21st century, a resurgence of interest in the testing effect has arisen, leading to the publication of many studies that explored various theoretical explanations for the phenomenon as well as applications to educational contexts.

**Generalizability of the Testing Effect**

The testing effect is a robust phenomenon: The basic finding has been replicated over a hundred times and its generalizability is well established. Retrieval practice has been found to promote superior retention of many different types of information, both verbal and nonverbal. These types of information include nonsense syllables, word lists, foreign language vocabulary, general knowledge facts, scientific articles, textbook chapters, pictures, maps, and Chinese characters (among others). In addition, many studies have shown strong, positive effects of testing in a variety of real-world educational contexts, such as after-school programs for elementary school children, middle school classes, college courses, and medical education of residents and nurses. Although for practical reasons, most testing effect studies have used relatively short retention intervals (i.e., a few minutes to a few days), a number of studies have shown that testing produces superior long-term
Retention using much longer retention intervals of up to six months. Overall, much evidence exists to support the conclusion that retrieval practice promotes long-term retention of many different types of materials across a variety of different contexts.

Potential Theoretical Explanations
Several theoretical explanations have been proposed to account for the testing effect. One of the first proposed that taking a test after studying resulted in additional exposure to the material (i.e., relative to a control condition in which no test was taken) and this additional exposure produced the superior retention. However, this so-called total time hypothesis (also referred to as the amount-of-processing hypothesis) was disproved by subsequent studies that showed that taking a test led to better retention relative to restudying the material for an equivalent amount of time. Another possible explanation is that the effort involved in retrieval is responsible for the testing effect. One piece of evidence that supports the retrieval effort hypothesis is the finding that production tests (e.g., cued recall, fill-in-the-blank), which require greater retrieval effort, often produce better retention than recognition tests (e.g., multiple-choice, true or false). Yet another idea that helps explain the testing effect is transfer-appropriate processing, which holds that memory performance is enhanced to the extent that the processes during encoding match the processes required during retrieval. Thus, retrieving information while taking an initial test may lead to better performance on a subsequent test because the processes engaged on an initial test (i.e., retrieval practice) better match the processes required by the final test (relative to restudying or not taking a test). Finally, the idea of encoding variability provides one more possible explanation. Studying and taking a test represent distinct encoding events, and thus testing after studying may increase encoding variability. Increased encoding variability should result in the elaboration of the existing memory trace and/or the creation of additional retrieval routes to that trace. Although no single theory can explain all the extant findings, the last four theories are not mutually exclusive and can be considered complementary.

Factors That Increase the Efficacy of Retrieval Practice
The critical mechanism in learning from tests is successful retrieval. However, two other factors can increase the efficacy of testing: feedback and repetition. Testing often produces better retention than restudying even when feedback is not provided (provided performance on the initial test is reasonably high). Nevertheless, feedback can enhance learning from tests by enabling test takers to correct errors and maintain low-confidence correct responses, thereby increasing the probability of successful retrieval in the future. Repetition can also enhance learning from tests: A single test confers a substantial mnemonic benefit, but repeated testing leads to even better retention. Repeated testing is particularly effective if it is distributed or spaced out over time rather than massed together. Generally speaking, spaced practice usually leads to superior long-term retention relative to massed practice, a finding that has been termed the spacing effect.

Henry L. Roediger III and Andrew C. Butler

See also Desirable Difficulties Perspective on Learning; Rehearsal and Memory; Spacing Effect; Spacing Effect, Practical Applications

Further Readings
Schizophrenia is a severe, chronic, psychiatric syndrome affecting about 1% of the world's population. In 1911, psychiatrist Paul Eugen Bleuler first coined the term, and since then, a large body of research has been amassed. Research has primarily focused on better understanding the etiology and progression of schizophrenia, while attempting to improve the lives of those with the disorder. As discussed in this entry, contemporary views of schizophrenia involve a syndrome with characteristic clusters of symptoms very similar to those first described by Bleuler and attributes those symptoms to a neurodevelopmental process, by which early and late risk factors contribute to the onset and expression of the disorder. The entry then examines cognitive and neurobehavioral impairments observed in the disorder as well as approaches to treatment.

Clinical Symptomatology of Schizophrenia

Characteristic clinical features of schizophrenia are classified into positive, negative, and disorganized symptoms. Positive symptoms include hallucinations (e.g., hearing voices that others cannot) and delusions (e.g., a persecutory delusion, involving the belief that others intend to harm the individual). Individuals with a diagnosis of schizophrenia also exhibit negative symptoms, which are defined by profound disruption of emotional expression and/or experience and motivation, often resulting in social withdrawal and a drop in day-to-day functioning.

The third category, disorganization symptoms, refers to bizarre behavior, tangential and disorganized speech, and illogical thought patterns. Schizophrenia is commonly associated with a heterogeneous presentation, with individuals exhibiting different combinations of these symptoms. In addition, a wide range of outcomes is observed. For example, some individuals show improved functioning between episodes of psychosis, whereas others display a more chronic course, with the continued presence of one or more of the above symptoms.

Neurodevelopmental Model of Schizophrenia

In its earliest clinical descriptions, schizophrenia was considered a deteriorating brain disorder with a course similar to Alzheimer's disease but an onset in young adulthood. Thus, the neurodegenerative hypothesis of schizophrenia initially prevailed. However, after nearly 3 decades of research, the neurodevelopmental model of schizophrenia is now prominent. This theoretical framework holds that the disorder's neural origins arise primarily during early development, with full emergence of recognizable symptoms typically occurring during late adolescence or early adulthood. By identifying associations between prenatal and perinatal complications and elevated risk for schizophrenia, studies have implicated that adverse events during early life may contribute to the development of the disorder. Furthermore, longitudinal studies have found that subtle deficits in cognition, emotional expression, and behavior are present during early childhood among individuals who ultimately develop...
schizophrenia as adults. These findings suggest that signs of brain compromise are present long before illness onset. Finally, the majority of postmortem neuropathology studies have failed to detect evidence of a neuronal degenerative process in schizophrenia. Although debate persists regarding the details of this aberrant neurodevelopmental course, when taken together, these findings provide compelling evidence that processes comprising neuronal development are fundamental to the pathophysiology of schizophrenia.

Early Risk Factors

Schizophrenia is a highly heritable illness, with approximately 80% of an individual's likelihood of developing schizophrenia attributable to their genetic makeup. There are now numerous studies demonstrating that unaffected biological relatives of patients with schizophrenia display qualitatively similar, but quantitatively milder, neuropsychological and neuroanatomic deficits, relative to healthy controls. Such findings support the view that these deficits reflect a genetic origin rather than secondary effects of the disease process or chronic medication use. Despite this strong genetic component, however, efforts to identify the precise risk genes involved have been challenging. In fact, not only is our understanding of which genes may be involved constantly being revised, so is our set of possible routes by which any given gene may result in phenotypic variation.

For instance, a major shift in our understanding of schizophrenia genetics comes from new studies of chromosomal structural variation, which indicate that rare mutations (copy number variants, or CNVs) may play a greater role in the etiology of schizophrenia than previously realized. CNVs likely comprise about 12% of the human genome and may be highly relevant for the expression of complex diseases. One particularly compelling example is the 22q11.2 deletion syndrome (22qDS), which represents the greatest known recurrent genetic risk factor to date for the development of schizophrenia. The cause of 22qDS is a deletion of one of an individual’s two copies of a particular section of chromosome 22q11.2, a locus that encompasses approximately 40 genes, including some known to play a key role in brain development. A well-defined neurogenetic syndrome like 22qDS can serve as a compelling model to help us understand how abnormal neurodevelopmental processes, which lead to brain dysfunction, can manifest in disturbances such as the clinical symptoms associated with schizophrenia. Furthermore, it suggests that while schizophrenia is a relatively common illness, it may represent the end product of a number of distinct developmental pathways.

While genetics are clearly a major risk factor, the fact that identical twins only show about 50% concordance for the illness (i.e., that both twins have a diagnosis of schizophrenia) indicates that environmental factors must also play a role. Prenatal and perinatal complications, particularly those associated with fetal hypoxia, or transient oxygen deprivation, appear to be among the environmental factors most robustly associated with increased risk. Furthermore, fetal oxygen deprivation offers a plausible mechanism for much of the structural brain pathology—such as hippocampal volume reduction—detected in neuroimaging studies of adult patients with schizophrenia. Other forms of pregnancy complications, such as prenatal viral exposure, also are associated with increased disease risk but to a lesser extent.

Late Risk Factors

Nevertheless, if risk factors for schizophrenia are at work during early brain development, why is it that the formal diagnostic symptoms and signs of the disorder do not typically manifest until late adolescence and early adulthood? The remarkable consistency in age-at-onset distributions for schizophrenia from around the world implicates late adolescence /early adulthood as the peak period of risk. According to the neurodevelopmental hypothesis of schizophrenia, the typical onset during late adolescence is likely related to widespread brain maturational changes occurring during this time period. As the typical brain matures through adolescence, there is an increase in prefrontal cortical white matter, which is crucial for efficient information transmission between brain regions. Concomitantly, there is a decrease in gray matter volume, likely as a result of synaptic pruning processes, which reduces the overall number of neurons in the brain, thereby retaining more efficient neural configurations.

Evidence from magnetic resonance imaging (MRI) studies suggests that those who go on to
develop psychosis show an exaggerated pattern of gray matter loss during this period. In comparison to those who do not develop a psychotic disorder, individuals who convert to psychosis show differential volume decreases in superior temporal and prefrontal regions, cortical regions important for high-order cognitive functions (e.g., planning, memory).

While the reasons for this dysregulation of typical brain maturation are still unknown, diathesis-stress models of schizophrenia suggest that environmental stressors may interact with genetic vulnerability in triggering symptom onset. While life stress is believed to exacerbate psychiatric symptoms across a variety of mental disorders, its role in precipitating illness onset in schizophrenia is controversial. Given the role of hormones in mediating stress response and in adolescent brain maturation, one possible biological mechanism by which rising hormone levels in adolescence may trigger expression of a latent genetic predisposition to schizophrenia is through dysregulation of the hypothalamic-pituitary-adrenal (HPA) axis. However, longitudinal studies are needed to determine causal relationships between early environmental insults, gonadal hormone expression, and the effects of chronic stress on HPA axis functioning and symptom expression in schizophrenia.

Neurobiological changes are likely to map onto behaviors that change during adolescence, such as social cognition. Social cognition, which broadly refers to the mental processes used to recognize, interpret, and respond to others’ social behavior, has been highlighted as one area of particular importance in schizophrenia. Deficits in emotion processing, the capacity to identify or discriminate between different emotions, and theory of mind (ToM), which refers to one’s ability to comprehend the intentions of others, are present prior to illness onset and appear to be relatively stable across phases of illness.

Because of the major changes that occur in one’s social environment during adolescence, it is likely that the development of social cognition in at-risk youth is detrimentally affected during this time. Efforts to understand this hypothesis are currently underway, particularly given an increasing emphasis on early identification and intervention for schizophrenia. The burgeoning area of clinical high risk (CHR) research aims to ascertain individuals initially showing symptoms indicating high risk for imminent onset of psychosis and follow them over time in order to characterize the course of neurobiological change among those who develop full-blown psychosis and to elucidate predictors of this outcome. These studies offer the unique opportunity to identify risk markers most predictive of schizophrenia outcome and, by extension, to develop interventions that may be implemented prior to onset of the full-blown disorder.

Results to date from CHR studies have identified several baseline clinical variables that appear to be predictive of conversion to psychosis over and above high-risk criteria alone. Approximately 30% of individuals considered at high risk for developing psychosis develop a full-blown psychotic disorder within 1 to 2 years after ascertainment. In the largest longitudinal study of CHR youth to date, poorer social functioning and a history of substance abuse increased one’s likelihood of conversion to psychosis. Neurocognitive studies have additionally identified verbal memory, processing speed, and working memory deficits as significant predictors of psychosis outcome.

Pathology of Schizophrenia

Developmental changes in brain structure may contribute to cognitive deficits seen in patients with established schizophrenia. Though subtle cognitive deficits are present long before the development of overt psychotic symptoms, it is likely that additional cognitive decline occurs right before or at illness onset. In schizophrenia patients, robust cognitive deficits have been observed in a variety of domains, including working and declarative memory, processing speed, and language production. It has been hypothesized that the wide array of observed cognitive impairments reflects global brain dysfunction or “dysconnectivity.”

Substantial evidence for cortical dysconnectivity in schizophrenia comes from electroencephalography (EEG), a means of examining the fluctuations in electrical field activity generated by synchronized activity of thousands of neurons. Disrupted synchronization of neurons results in abnormal electrical activity as well as disruption in the information-processing tasks associated with the observed brain activity. Decades of EEG research have demonstrated that patients with schizophrenia process information atypically, even when information-processing demands are very rapid and patients’ overt behavior
does not appear to be abnormal. For example, measurement of the earliest electrical oscillations measureable within 200 milliseconds after visual stimuli appear has shown that individuals with schizophrenia show an impaired neural response selectively to low-contrast objects, despite normal performance on a routine vision examination. A similar pattern of findings applies to rapid changes in auditory information, and dramatic abnormalities persist during the assessment of more complex cognitive tasks as well. These diverse, characteristic aberrations in the brain’s electrical activity are all thought to be signs that the coordinated activity within and between neural circuits is disrupted.

As is the case with behavioral and MRI-based measures, there is rapid change in large-scale electrical brain activity right before and immediately after onset of full symptoms of schizophrenia, which then remains relatively stable over the course of illness.

Treatment of Schizophrenia

Currently, the first line of treatment in schizophrenia is pharmacological intervention, which typically involves administration of antipsychotic medication. First-generation, or typical, antipsychotics (e.g., chlorpromazine, haloperidol) were first introduced in the 1950s. These drugs are thought to work by blocking dopamine receptors, and they are effective in decreasing the severity of positive psychotic symptoms, particularly hallucinations. However, first-generation antipsychotics carry significant side effects, particularly extrapyramidal motor symptoms (e.g., rigid body tremors). In the 1990s, second-generation or “atypical” antipsychotics (e.g., risperidone, ziprasidone, olanzapine) were introduced as an alternative; these medications were associated with fewer extrapyramidal symptoms than the typical antipsychotics. However, it is controversial whether atypical antipsychotics are actually safer; side effects associated with atypical antipsychotics include weight gain and increased risk for diabetes and stroke. Furthermore, although antipsychotic medications attenuate the presence of positive symptoms, the negative symptoms and cognitive dysfunction associated with schizophrenia often remain and continue to substantially impact functioning.

Given these challenges, many psychosocial interventions have been developed for schizophrenia. Early results indicate that cognitive-behavioral therapy, which focuses on learning to implement more adaptive behavioral responses to one’s thoughts and emotions, has shown to reduce positive and negative symptoms of schizophrenia and prevent relapse. Studies have also examined the effectiveness of social skills training programs, which are designed to address the social cognition dysfunction seen in schizophrenia. These studies have shown that after participating in a social skills training program, individuals with schizophrenia show improvements in their ability to recognize emotions and report having more social relationships. Many other psychosocial interventions, including vocational training and family psychoeducation, have been established and have demonstrated improved outcomes in schizophrenia. However, perhaps most important, research has shown that these psychosocial interventions are only effective when used in conjunction with medication and that timing of both pharmacological and psychosocial intervention is key, with reduced severity in the course of illness seen in earlier intervention.

Conclusions

Schizophrenia is a chronic and highly debilitating brain disorder. Currently, available treatments are palliative in nature; thus, there is an increasing emphasis in the field on early identification and early intervention. The case for involvement of early neurodevelopmental influences in the pathogenesis of schizophrenia is compelling, including evidence from human epidemiological studies, prospective studies of birth cohorts and at-risk populations, and postmortem neuropathology studies. The case for involvement of later neurodevelopmental processes is still largely circumstantial, but abnormalities of synaptic pruning processes during adolescence are likely relevant for symptom onset. A better understanding of the developmental trajectory of neurobiological processes in schizophrenia will inform early intervention strategies as to the most vulnerable brain structures and functions, as well as the stages of the illness most amenable to treatment.

Maria Jalbrzikowski, Peter Bachman, and Carrie E. Bearden

See also Behavioral Therapy; Delusions; Emotion and Psychopathology; Social Cognition
Further Readings


**Scientific Reasoning**

Scientific reasoning, like science itself, is a constantly changing and inherently fascinating system of interrelated concepts, practices, and theoretical approaches to conducting science. This entry discusses the philosophical and empirical approaches used to understand the scientific process and how those approaches have coevolved with it. Early research on scientific thinking offered conclusions that were potentially universal for all sciences about the formation of hypotheses, optimizing scientific research, and the implications of these conclusions for science education. Modern research on scientific thinking has become much more specific to individual scientific disciplines (such as molecular biology), and it studies the way science is conducted in the real world by observing scientists as they work or by analyzing their diaries and research notes. The new field of educational neuroscience has outlined the brain structures involved in such reasoning as well. This entry also explains the importance of viewing science in its appropriate historical context and examines how modern scientific thinking has been influenced by the vast capabilities of robotics and computers.

Just 20 years ago the field was dominated by a small set of questions regarding the best ways to conduct science, the relationship between hypotheses and experiments, what scientific reasoning strategies should be taught, and whether there could be a unified science of science. Much cognitive and philosophical work focused on when and whether scientists should attempt to confirm or disconfirm their hypotheses. Many researchers followed the view of Karl Popper that scientists should attempt to disconfirm their hypotheses. However, more fine-grained research demonstrated that many scientists seek to confirm their hypotheses early on in a research project and seek disconfirmation at later stages of research. Researchers also focused on whether science is primarily inductive (inferring general rules from a finite number of observations), deductive (generating specific conclusions from known, general rules), or abductive (inferring a cause that would best explain a given effect) and found that these different forms of reasoning are all used in science, rather than science being one form of reasoning exclusively.

Understanding Science by Modeling Real Life Situations

The reasoning strategies of renowned scientists and their discoveries have been examined in historical analyses, often using computer simulations of scientific discovery. Historical and computational approaches have revealed that specific scientific reasoning strategies such as following up unexpected results, using analogies to formulate hypotheses, and assessing the coherence of a scientific concept are key features of scientific thinking. Furthermore, investigations of students reasoning scientifically and children reasoning about scientific concepts have demonstrated that search in different types of problem spaces is central to understanding the development of scientific thinking. A problem space...
includes the current state of knowledge, the goal state (which may not be defined), and all knowledge states in between, as well as the cognitive operators that allow one to move from one knowledge state to the next. Many researchers have adopted a more detailed approach to the development of scientific thinking and have moved away from Jean Piaget's stagelike view of the development of scientific thinking skills to an investigation of scientific thinking strategies that can be taught in the classroom, such as designing experiments, formulating hypotheses, and learning how to assess the adequacy of particular experimental designs.

**Putting Science in Its Historical Context**

Another strand in research on scientific thinking over the past 50 years has been to investigate the historical context of particular scientific concepts. The genesis of this view of scientific thinking was undoubtedly Thomas Kuhn's *Structure of Scientific Revolutions*, in which Kuhn demonstrated that science moves in ways reminiscent of a political revolution rather than by a steady accrual of knowledge. This realization that science is not the strict accumulation of logical facts, findings, and methods led to many socially grounded theories of scientific thinking. According to this approach, the adoption or abandonment of a scientific theory may be due to many factors and not just whether a theory explains a set of "facts." Consequently, many recent historical analyses of scientific thinking have demonstrated that conceptual change and theory change in science occur for a wide variety of both nonscientific and scientific reasons.

**Modern Approaches to Scientific Thinking**

Much research in science education and scientific thinking has concentrated on the mechanisms of conceptual change for specific scientific concepts. Research on scientific thinking has moved away from domain-general experiments (studies that can be applied to any number of different scientific disciplines) on how people test simple hypotheses, to more complex domains with complex histories such as molecular biology, physics, evolution, and chemistry. Researchers now routinely investigate the scientific thinking strategies that scientists themselves use in their own labs to understand both the cognitive and social factors involved in scientific thinking.

In the early 2000s, cognitive neuroscientists began to explore the neural underpinnings of scientific thinking, demonstrating the roles of the anterior cingulate and parahippocampal gyrus when students ignore evidence that disconfirms their favored hypothesis, the role of the frontal poles when generating scientific analogies, and the role of the left dorsolateral prefrontal cortex and precuneus in causal scientific thinking. This approach to the brain, education, and cognition has helped lead to the development of the new field of educational neuroscience, in which educationally important questions are addressed using a variety of populations, neuroimaging techniques, and educationally rich contexts to produce more effective teaching, robust learning, and in-depth understanding.

**A New Scientific Possibility: The Hypothesis-Optional Experiment**

Around the same time as the development of educational neuroscience, major changes began to take place in the ways that scientists use hypotheses and conduct experiments: Many biologists started conducting massive, catchall projects in which millions of experiments—essentially all possible experiments that could be conducted within a particular problem space—could be conducted in a short space of time by automating the process with robotics and supercomputers. These high throughput experiments (conducted by computational devices, robots, and machines) are now routine in the pharmaceutical and genetics industries, and many researchers have argued that this approach obviates the need for hypotheses at all. If this approach endures, it fundamentally changes the nature of scientific thinking. This radical shift in some of the basic assumptions of scientific thinking will be a central focus of research for the next few decades as more of the high throughput technologies are used in all branches of science ranging from astrophysics to zoology. This shift will undoubtedly lead to the proposal of many new types of scientific thinking and will change scientific practice as much as Sir Francis Bacon's introduction of inductive methods in the 1600s changed the nature of science in the subsequent centuries.

*Kevin N. Dunbar and Evelyn A. Forster*

*See also* Analogical Mapping and Reasoning; Deductive Reasoning; Distributed Cognition
Self, Development of

“Am I Me?” a thoughtful 2-year-old queried of his parents. Beginning in the second year of life, toddlers begin to talk about themselves. They master self-relevant personal pronouns (I and me) that distinguish themselves from others. With development, they come to understand that they possess various characteristics, some of which may be positive (“I’m smart”) and some of which may be negative (“I’m unpopular”). Of particular interest is how the nature of such self-evaluations changes with development as well as how self-evaluations differ among individual children and adolescents across two basic evaluative categories. The first category is (a) domain-specific self-concepts—namely, how one judges one’s attributes in particular arenas—for example, scholastic competence, social acceptance, physical competence, and so forth. A given individual may vary tremendously in how they feel across these domains, creating a meaningful profile of scores. One typically does not feel equally adequate across all domains. The second evaluative category is (b) global self-esteem—namely, how much they value their overall worth as a person. (For a complete treatment of self-development in childhood and adolescents, see Susan Harter’s work.)

Developmental shifts in the nature of self-evaluations are driven by changes in the child’s cognitive capabilities. Cognitive-developmental theory and findings by Jean Piaget and Kurt Fischer alert us to the fact that the young child is limited to very specific, concrete representations of self and others, for example, “I know my A, B, Cs.” In middle to later childhood, a child develops the ability to form higher order concepts about his or her attributes and abilities (e.g., “I’m smart”). There are further cognitive advances at adolescence, allowing the teenager to form abstract concepts about the self that transcend concrete behavioral manifestations and higher order generalizations (e.g., “I’m intelligent”).

Developmental Differences in Domain-Specific Self-Concepts

Domain-specific self-concepts are observed at every developmental level. However, the precise nature of these judgments varies with age. There are five common domains in which children and adolescents make evaluative judgments about the self: scholastic competence, physical competence, social competence, behavioral conduct, and physical appearance. The types of statements vary, however, across three age periods—early childhood, later childhood, and adolescence—in keeping with the cognitive abilities and limitations of each age period.

Early Childhood

Young children provide very concrete accounts of their capabilities, evaluating specific behaviors. Thus, they communicate how they can count, how they can run very fast, how they are nice to a particular friend, how they don’t hit their sister, and how they possess a specific physical feature such as pretty blond hair. Of particular interest in such accounts is the fact that the young child typically provides a litany of virtues, touting their positive skills and attributes. One cognitive limitation of this age period is that young children cannot distinguish the wish to be competent from reality. As a result, they typically overestimate their abilities because they do not yet have the skills to evaluate themselves realistically. Another cognitive characteristic that contributes to potential distortions is the pervasiveness of all-or-none thinking. That is, evaluations are either all positive or all negative. With regard to self-evaluations, they are typically all positive. (Exceptions to this

Further Readings


positivity bias can be observed in children who are chronically abused, since severe maltreatment is often accompanied by parental messages that make the child feel inadequate, incompetent, and unlovable. Such children will also engage in all-or-none thinking but conclude that they are all bad.

**Middle to Later Childhood**

As the child grows older, the ability to make higher order generalizations in evaluating his or her abilities and attributes emerges. Thus, in addition to describing his or her prowess at a particular activity, the child will also observe that he or she is good at sports, in general. This inference can further be justified in that the child can describe his or her talent at several sports (e.g., good at soccer, basketball, and baseball). Thus, the higher order generalization represents a more developmentally advanced cognitive construction in which an overarching evaluation (e.g., “I am good at sports”) is defined in terms of specific examples that justify this conclusion. Similar processes allow the older child to conclude that he or she is smart (e.g., does well in math, science, and history). The structure of a higher order generalization about being well behaved could include such components as obeying parents, not getting into trouble, and trying to do what is right. A generalization concerning one’s popularity may subsume accounts of having friends at school, making friends easily at camp, and developing friendships readily when moving into a new neighborhood. The perception that one is good looking may be based on one’s positive evaluation of one’s face, hair, and body.

During middle childhood, all-or-none thinking diminishes and the aura of positivity fades. Thus, children do not typically think that they are all virtuous in every domain. The more common pattern is for them to feel more adequate in some domains than others. For example, one child may feel that he or she is good at schoolwork and is well behaved, whereas he or she is not that good at sports, does not think that he or she is good-looking, and reports that it is hard to make friends. Another child may report the opposite pattern.

There are numerous combinations of positive and negative evaluations across these domains that children can and do report. Moreover, they may report both positive and negative judgments within a given domain—for example, they are smart in some school subjects (math and science) but dumb in others (English and social studies). Such evaluations may be accompanied by self-affects—namely, emotions about the self that emerge in later childhood—for example, feeling proud of one’s accomplishments but ashamed of one’s perceived failures, as Susan Harter’s work documents. The ability to consider both positive and negative characteristics is a major cognitive-developmental acquisition. Thus, beginning in middle to later childhood, these distinctions result in a profile of self-evaluations across domains.

Contributing to this advance is the ability to engage in social comparison. Beginning in middle childhood one can utilize comparisons with others as a barometer of the skills and attributes of the self. In contrast, the young child cannot simultaneously compare his or her attributes to the characteristics of another in order to detect similarities or differences that have implications for the self. Although the ability to utilize social comparison information for the purpose of self-evaluation represents a cognitive-developmental advance, it also ushers in new, potential liabilities. With the emergence of the ability to rank order the performance of other children, all but the most capable children will necessarily fall short of excellence. Thus, the very ability and penchant to compare oneself with others makes the self vulnerable, particularly if one does not measure up in domains that are highly valued. The more general effects of social comparison can be observed in findings revealing that domain-specific self-concepts become more negative during middle and later childhood compared to early childhood.

**Adolescence**

For the adolescent, there are further cognitive-developmental advances that alter the nature of domain-specific self-evaluations. Adolescence brings with it the ability to create more abstract judgments about one’s attributes and abilities. Thus, one no longer merely considers oneself to be good at sports but to be athletically talented. One is no longer merely smart but views the self more generally as intelligent, where successful academic performance, general problem-solving ability, and creativity might all be subsumed under the abstraction of intelligence. Abstractions may be similarly constructed in the other domains. For example, in the domain of behavioral conduct, there will be a shift from the
perception that one is well behaved to a sense that one is a moral or principled person. In the domains of social competence and appearance, abstractions may take the form of perceptions that one is popular and physically attractive.

These illustrative examples all represent positive self-evaluations. However, during adolescence (as well as in later childhood), judgments about one’s attributes will also involve negative self-evaluations. Thus, certain individuals may judge the self to be unattractive, unpopular, unprincipled, and so forth. Of particular interest is the fact that when abstractions emerge, the adolescent typically does not have total control over these new acquisitions, just as when one is acquiring a new athletic skill (e.g., swinging a bat, throwing a ball, maneuvering skis), one lacks a certain level of control. In the cognitive realm, such lack of control often leads to overgeneralizations that can shift dramatically across situations or time. For example, the adolescent may conclude at one point in time that he or she is exceedingly popular but then, in the face of a minor social rebuff, may conclude that he or she is extremely unpopular. This typically leads to exasperation by parents and other adults in the adolescent’s life since they do not understand that such shifts are inevitable and quite typical in our culture. The shifts stem from new cognitive advances that also reflect liabilities. Plus, recent work reveals that immature brain structures contribute, structures that preclude the more integrated thought that helps control such vacillations. Gradually, adolescents gain control over these self-relevant abstractions so that they become capable of more balanced and accurate self-evaluations, as research by Fischer and by Harter reveals.

Global Self-Esteem

The ability to evaluate one’s worth as a person and to make inferences about one’s self-esteem also undergoes developmental change. The young child simply is incapable, cognitively, of developing the verbal concept of his or her value as a person. This ability emerges at the approximate age of 8. However, young children exude a sense of personal worth in their behavior. The primary behavioral manifestations involve displays of confidence, independence, mastery attempts, and exploration, as Susan Harter’s work documents. Thus, behaviors that communicate to others that children are sure of themselves reflect high self-esteem in early childhood. In contrast, behavior indicative of lack of confidence, mastery attempts, curiosity, or exploration, plus excessive dependence on others, reflects a constellation that is predictive of low self-esteem in others’ eyes, such as teachers or parents.

At about the third grade, children begin to develop the concept that they like or don’t like the kind of person they are, as the writings of Susan Harter and Morris Rosenberg indicate. Thus, they can respond to general items asking them to rate the extent to which they are pleased with themselves, like who they are, and think they are fine as a person. Here, the shift reflects the emergence of an ability to construct a higher order generalization about the self. This type of concept can be built on perceptions that one has a number of specific qualities—for example, that one is competent, well behaved, attractive, and so forth. Self-esteem can also be built on the observation that significant others—for example, parents, peers, and teachers—think highly of themselves. This process is greatly influenced by advances in the child’s ability to take the perspective of significant others. During adolescence, one’s evaluation of one’s global worth as a person may be further elaborated, drawing on more domains and sources of approval, and will also become more abstract. Thus, adolescents can directly acknowledge that they have high or low self-esteem as a general abstraction about the self; that is, they understand the term self-esteem.

Individual Differences in Domain-Specific Self-Concepts as Well as Global Self-Esteem

Although there are predictable cognitively based developmental changes in the nature of how most children and adolescents describe and evaluate themselves, there are striking individual differences in how positively or negatively the self is evaluated. Moreover, one observes different profiles of children’s perceptions of their competence or adequacy across the various self-concept domains, in that children evaluate themselves differently across domains.

Consider two profiles exemplified by Child A and Child B, neither of whom feels good about the self scholastically or athletically. They evaluate themselves much more positively in the domains of social acceptance, conduct, and physical appearance. In fact, their profiles are quite similar to each
other across the five specific domains. However, judgments of their self-esteem are extremely different. Child A has very high self-esteem whereas Child B has very low self-esteem. This raises a puzzling question: How can two children look so similar with regard to their domain-specific self-concepts but evaluate their global self-esteem so differently? We turn to this issue next, in examining the causes of global self-esteem.

The Causes of Children’s Level of Self-Esteem

Our understanding of the antecedents of global self-esteem has been greatly aided by the formulations of two historical scholars of the self, William James and Charles Horton Cooley. Each suggested rather different pathways to self-esteem, defined as an overall evaluation of one’s worth as a person. William James focused on how the individual assessed his or her competence in domains where one had aspirations to succeed. Charles Horton Cooley focused on the salience of the opinions that others held about the self, opinions that one incorporated into one’s global sense of self.

Competence/Adequacy in Domains of Importance

For William James, global self-esteem derived from the evaluations of one’s sense of competence or adequacy in the various domains of one’s life relative to how important it was to be successful in these domains. Thus, if one feels one is successful in domains deemed important, high self-esteem will result. Conversely, if one falls short of one’s goal in domains where one has aspirations to be successful, one will experience low self-esteem. One does not, therefore, have to be a superstar in every domain to have high self-esteem. Rather, one only needs to feel adequate or competent in those areas judged to be important to the self. Thus, children may evaluate themselves as unathletic; however, if athletic prowess is not an aspiration, then self-esteem will not be negatively affected. That is, the high self-esteem individual can discount the importance of areas in which one does not feel successful.

This analysis can be applied to the profiles of Child A and Child B. In fact, Susan Harter has directly examined this explanation by asking children to rate how important it is for them to be successful. The findings reveal that high self-esteem individuals feel competent in domains they rate as important. Low self-esteem individuals report that areas in which they are unsuccessful are still very important to them. Thus, Child A represents an example of an individual who feels that social acceptance, conduct, and appearance, domains in which she evaluates herself positively, are very important but that the two domains where she is less successful, scholastic competence and athletic competence, are not that important. In contrast, Child B rates all domains as important, including the two domains where he is not successful, scholastic competence and athletic competence. Thus, the discrepancy between high importance coupled with perceptions of inadequacy contributes to low self-esteem.

Incorporation of the Opinions of Significant Others

Another important factor influencing self-esteem can be derived from the writings of Charles Horton Cooley, who metaphorically made reference to the “looking-glass self.” According to this formulation, significant others (e.g., parents and peers) are social mirrors into which one gazes to determine their opinion of the self. Thus, in evaluating the self, one would adopt what one felt were the judgments of these others whose opinions are considered important. Thus, the approval, support, or positive regard from significant others becomes a critical source of one’s own sense of worth as a person. For example, children who receive approval from parents and peers will report much higher self-esteem than children who experience disapproval from parents and peers.

Findings reveal that both of these factors, competence in domains of importance and the perceived support of significant others, combine to influence a child’s or adolescent’s self-esteem. Those who feel competent in domains of importance and who report high support rate themselves as having the highest self-esteem. Those who feel inadequate in domains deemed important and who report low levels of support rate themselves as having the lowest self-esteem. Other combinations fall in between.

Conclusions

Two types of self-evaluations that can be observed in children and adolescents were distinguished:
(a) evaluative judgments of competence or adequacy in specific domains (domain-specific self-concepts) and (b) the global evaluation of one's worth as a person—namely, overall self-esteem. Each of these undergoes developmental change based on age-related cognitive advances. In addition, older children and adolescents vary tremendously with regard to whether self-evaluations are positive or negative. Within a given individual, there will be a profile of self-evaluations, some of which are more positive and some that are more negative. More positive self-concepts in domains considered important as well as approval from significant others will lead to high self-esteem. Conversely, negative self-concepts in domains considered important coupled with lack of approval from significant others will result in low self-esteem.

Why Should We Care About Self-Concepts and Self-Esteem?

Self-esteem is particularly important since it is associated with very critical outcomes or consequences. Perhaps the most well-documented consequence of low self-esteem is depression. Children and adolescents (as well as adults) who display the constellation of low perceived adequacy in domains of importance coupled with low approval support from significant others invariably report low self-esteem. Low self-esteem, in turn, is highly associated with perceived depression and hopelessness about the future. The most seriously depressed consider suicide. Thus, it is critical that we intervene for those experiencing low self-esteem. Our model of the causes of self-esteem suggests strategies that may be fruitful—for example, improving skills, helping individuals discount the importance of domains in which it is unlikely that they can improve, and providing support in the form of approval for who they are as people. It is also clear that there are different pathways to low and high self-esteem. For example, for one child, the sense of inadequacy in particular domains may be the pathway to low self-esteem. For another child, lack of support from parents or peers may represent the primary cause. These different pathways are important to identify because they have critical implications for intervention efforts to enhance feelings of worth for those children with low self-esteem in educational settings, clinical settings, peer groups, the community, and within the family. Susan Harter’s work delineates multiple strategies for promoting realistically high self-esteem. If evaluations are accurate, positive self-esteem is clearly a psychological commodity, a resource that is important for us to foster in our children and adolescents if we want them to lead productive and happy lives.

Susan Harter

See also Concepts, Development of; Personal Identity, Development of; Relationships, Development of

Further Readings

Self-Consciousness

The term self-consciousness has several often conflated usages in the philosophy of mind and the cognitive sciences. This entry distinguishes the three major common usages and their general philosophical and scientific contexts, with an emphasis on their place in cognitive neuroscience.
Three Usages of Self-Consciousness

1. Core self-consciousness. According to this usage, self-consciousness is an essential structural property of consciousness that conditions all other forms of awareness. The controversial claim presupposed by the usage (sometimes referred to as the “self-awareness thesis” or SAT) is that all conscious beings possess this type of self-consciousness irrespective of their conceptual sophistication or their capacities for introspection. According to the SAT, consciousness is necessarily aware of itself in all its streams or episodes. The SAT has been held, in one form or another, over millennia by a variety of philosophers, psychologists, neuroscientists, and other investigators of consciousness, including, notably, Aristotle, René Descartes, John Locke, Immanuel Kant, Franz Brentano, Sigmund Freud, Edmund Husserl, and the phenomenologists who followed Husserl. The SAT has had few supporters in Anglo-American philosophy of mind, but this has recently changed. In the cognitive neuroscience literature, the most notable proponent of the SAT is Antonio Damasio, with his distinction between core and extended consciousness and his focus on the primordial experience of the body and emotion. In the artificial intelligence-inspired literature, it is Douglas Hofstadter. The qualifier “core” derives from the work of Damasio; in the literature, one will find many qualifiers used to designate this form of self-consciousness (e.g., pre-reflective, nonpositional, nonthetic, marginal, inattentive, peripheral, tacit).

2. Introspective self-consciousness. The second usage refers to the exercise of the ability to attend to, conceptualize, and report on one’s mental states. This form of self-consciousness is also sometimes marked by the qualifiers “reflective” and “attentive.” This form of self-consciousness plays a crucial role in normal mental life: in self-orienting and ongoing inner dialogue, in reconceptualization of personality traits and goals, and in sharing intentions, feelings, and so forth with others. Presumably, animals and infants could be self-conscious in the first sense of the term but have limited or no introspective self-consciousness. Introspective self-consciousness plays a critical role in the practice of cognitive psychology and neuroscience, both in the scientists’ own heuristic introspection and in the self-reports of experimental subjects, a crucial source of data in many experimental paradigms. However, it remains subject to multiple sources: “noise” fatigue, distraction, failure to communicate or understand instructions, personality and emotional biases, social conformity, and confabulation. Researchers have called for the development of rigorous experimental methods of cross validation to minimize errors stemming from such limitations and to allow for the better correlation of ongoing experiences with spontaneous brain activity.

3. Extended self-consciousness. The third usage is sometimes indicated with the qualifier “autobiographical,” and refers to the conscious access to and projection of the memory-laden information necessary for conceptualizing and situating oneself as an ego, self, person, or responsible agent with a culturally and socially mediated history and anticipated future. It is closely related to the notion of the “self-concept.” Extended self-consciousness may be the most derivative of the three as it seems to depend on the former two. It requires an extended degree of conceptualization of space, time, dispositions, causal relations, and moral notions, and is likely most developed in adults. Its accuracy depends on the time and resource-consuming cognitive ability to reflectively notice and integrate one's stream of experiences, patterns of behaviors, affective reactions, interests, and motivations. It can be biased by various motives, personality traits, and coping mechanisms. The self-concept tends to undergo development during childhood, adolescence, and midlife and can be modified by positive or negative experiences (e.g., love, loss, depression, and treatment). Finally, the self-concept can be eroded by degenerative disease (e.g., Alzheimer’s disease [AD]) or strongly affected by other forms of brain damage, while the subject can remain otherwise self-conscious in senses one and two. Extended self-consciousness is thus a far more dynamic property than the other two types. During the progression of AD, extended self-consciousness is among the first cognitive functions to be affected, followed by introspective self-consciousness, while core self-consciousness is among the last functions to disappear. Core, introspective, and even some extended self-consciousness appear to remain largely preserved in neuropsychological cases of patients with profound retrograde and anterograde amnesia.
Many authors cite experiences of complete absorption during activities in which one seems to lose track of oneself as evidence that there can be consciousness without any self-consciousness whatsoever. Beyond the potential paradox of such a position, which seems to imply the possibility of experiences (e.g., pain) with no owner or subject of the experience, it seems generally to assume that self-consciousness is to be taken in senses two or three. While it is true that one is not introspecting all the time and that one ceases to think about oneself as an autobiographical ego in times of complete absorption, it also does not seem entirely implausible that animals, infants, and sufferers of degenerative disease could be self-conscious without introspective capacity and without having much of an autobiographical self.

Neural Bases of Self-Consciousness

The study of the neural basis of self-consciousness has grown over the past decade. Three classes of hypotheses can be distinguished.

1. The first class refers to specific regions of the brain as underlying self-consciousness or its components, including the following: (a) the insular cortex—for all forms of self-consciousness; (b) the anterior cingulate cortex—for interoceptive and emotional self-consciousness, the conscious monitoring of conflict, and introspective self-consciousness; (c) the medial prefrontal cortex—for self-referential processing in general, and (d) the brain stem, which has recently been hypothesized to be the neural basis for the “primordial feelings” of the living body.

2. The second class of hypotheses focus on more distributed and integrated networks, including the following: (a) the default mode network that has been hypothesized to constitute an essential basis for introspection, the retrieval of autobiographical memory, and the projection of the self into the future; (b) a network of subcortical-cortical midline structures in which cortical midline components subserve introspective self-consciousness and core self-consciousness is implemented by subcortical components; (c) a posterior network relying on interactions between brain stem nuclei, the thalamus, and the posteromedial cortex, which has been hypothesized to constitute an integrative basis for “subjectivity”—that is, essentially core and introspective self-consciousness.

3. A third class of hypotheses considers self-consciousness as potentially “multiply realizable.” It would function like the implementation of an abstractly characterizable computational algorithm capable of exploiting different brain regions in an equipotent manner, similar to the way a computer’s virtual machine can flexibly occupy variable hardware resources.

Compatible with the second and third classes of hypotheses, neurological evidence supports the idea that self-consciousness is an extremely robust, flexible, and resilient process. Neurological studies suggest that, beyond coma caused by brain stem and thalamic damage, self-consciousness may disappear only when extensive bilateral damage affects most of the thalamocortical system. Even in such circumstances, as in some persistent vegetative state patients, there is evidence that residual self-consciousness can be present. Moreover, while genetic and environmental factors can result in aberrant brain development, there is often a preservation of basic mental abilities, including self-consciousness. Developmental cases suggest that extremely different brains across individuals can support similar forms of self-consciousness.

Understanding the structure and mechanisms underlying core self-consciousness will represent a key theoretical and methodological challenge for the cognitive sciences. Likewise, uncovering the neural implementation of self-consciousness will represent a tremendous challenge if the “multiple realizability” hypothesis turns out to be true, as the standard anatomo-functional approach of cognitive neuroscience would be of limited help.

Kenneth Williford, David Rudrauf, and Carissa L. Philippi

See also Anosognosia; Consciousness, Comparative Perspectives; Event Memory, Development; Introspection; Self, Development of; Self-Knowledge

Further Readings

Self-Knowledge

Self-knowledge is the characteristically human ability of knowing one’s own mental states—such as sensations, perceptions, emotions, and propositional attitudes. The following will survey the main philosophical and psychological accounts of self-knowledge proposed in recent years.

The Introspective Model

According to the Cartesian conception, all mental states are like objects presented in one’s own mental arena we are introspectively aware of. In particular, they are transparent to their subjects; that is, if a subject has them, he is immediately aware of them and in a position to judge that he has them. Moreover, a subject is authoritative with respect to them—if sincere, if he judges to be in a mental state M, he is.

The Cartesian model has been widely criticized. First, since Sigmund Freud’s discovery of the unconscious, the transparency of mental states has been questioned. Second, it has been noted that animals and infants have mental states yet can’t self-ascribe them. Third, the discovery of self-deception, whereby subjects self-ascribe mental states they don’t actually have, makes authority founder. Fourth, the Cartesian model would introduce a cognitive faculty—namely, introspection—modeled after sight, which, however, appears difficult to characterize in relation to mental states that aren’t physical entities. Finally, Ludwig Wittgenstein pointed out how the Cartesian model would entail the view that psychological language is private to each subject. On the one hand, one could know only one’s own mental states and may merely surmise those of others. On the other, if only I can know my current mental states, the reference of my psychological vocabulary will be known to myself only. Hence, whenever it will seem correct to me to apply a given term t to a current mental state M, my use of t will be correct. According to Wittgenstein’s argument against the possibility of a private language (either in speech or in thought), this would entail that the distinction between correct and incorrect uses of t would collapse and, with it, the very idea that t could mean anything at all.

David Armstrong has proposed a refined version of the introspective model. In this view, through the operation of an inner subpersonal mechanism, subjects would become immediately aware of their mental states. Working reliably, such a mechanism would also ensure that subjects be authoritative about them.

Armstrong’s model has been criticized for considering both transparency and authority to be the result of the correct operation of the inner subpersonal mechanism. These characteristics should be subject to perfectly acceptable exceptions in case the mechanism broke down. Authority and transparency, however, are traditionally considered constitutive traits of self-knowledge and their failure either implies a lack of conceptual competence or rationality or else is a sign that the mental state one fails to have knowledge of is unconscious. Yet—to
contrast this view with how we usually think of subpersonal mechanisms—we would never think of blind subjects, whose visual mechanism is impaired and can’t therefore see objects presented to them, that they either lack the relevant concepts or are being irrational or else are unconsciously seeing the objects.

The Inferential Model

The inferential model, somewhat traceable to behaviorism, has been proposed by the psychologist Alison Gopnik as part of her theory theory conception of knowledge of the mind. Accordingly, subjects at the ages of 3 to 4 would know their own mental states in the same way as they know other people’s—namely, by making inferences from their overt behavior to their likely mental causes. Transparency and authority, therefore, aren’t constitutive features of self-knowledge but are mere illusions. “Authority” is due to the fact that, being continuously around ourselves, we become reliable at recognizing our own mental states. The illusion of transparency, in contrast, can be explained by means of an analogy: Just as trained physicists can immediately see electrons in a cloud chamber because they have acquired a theory and can immediately and reliably apply it, so human beings can self-ascribe mental states without going through explicit inferences because they quickly and reliably apply their own theory of mind.

The theory theory account of self-knowledge has been criticized on various grounds. Like Armstrong’s model, this account would turn transparency and authority into contingent rather than constitutive features of self-knowledge. In addition, it would deny the intuitive asymmetry between knowledge of our own and of other minds.

Constitutive Accounts

Constitutive accounts, developed by philosophers such as Sydney Shoemaker, Crispin Wright, and Akeel Bilgrami, reject the view that self-knowledge be based either on introspection or on inference. Furthermore, they consider transparency and authority to be constitutive features of self-knowledge. They therefore turn them into two conceptual truths, captured by the following thesis:

Given conditions C, a subject believes/judges that he is in mental state M if and only if he is.

Conditions C restrict the universal applicability of the thesis so as to counter the objections already raised against the introspective account of transparency and authority—that is, the arguments from unconscious mental states, from the mental states of animals and infants, and from self-deception. This restricted thesis holds only for subjects who are conceptually endowed and for specific kinds of mental states, like conscious beliefs, desires, and intentions. Moreover, according to Bilgrami, self-deceived subjects do not make wrong self-ascriptions but rather have two contrasting mental states—only one of which is conscious and correctly ascribed—that explain their conflicting and irrational behavior. Hence, the authority over their self-ascriptions is compatible with their being self-deceived.

Constitutive accounts have been criticized for failing to accommodate two intuitions: that first- and second-order mental states—that is, beliefs, desires, and intentions, on the one hand, and our own beliefs about them, on the other—have separate existence, and that self-knowledge is due to some sort of cognitive accomplishment.

Recent Epistemic Accounts

Richard Moran and Christopher Peacocke have claimed that self-knowledge consists in making judgments about one’s own beliefs, desires, and intentions on the basis of having them and for the reason that one has them. Self-knowledge is thus a modest yet genuinely cognitive accomplishment consisting in immediate, nonobservational judgment about one’s own mental states, rationally grounded in their obtaining.

Epistemic accounts, however, are problematic. In particular, awareness of one’s beliefs, desires, and intentions must avoid presupposing the very knowledge of them, which should be explained and yet be such that their self-ascription be rational—that is to say, consciously motivated by the occurrence of these mental states rather than simply caused by them. As a matter of fact, however, it is very difficult to see what conception of awareness of one’s own mental states could serve this purpose. For, on the one hand, so-called phenomenal awareness—that is, awareness of what it is like to have a given belief, desire, or intention—would arguably fail to provide subjects with reasons for their self-ascriptions. “Propositional” (or “higher order”) awareness, on
the other, would in fact presuppose self-knowledge as it would consist in judging of being in the relevant mental states.

Annalisa Coliva

See also Behaviorism; Belief and Judgment; Consciousness and the Unconscious; Folk Psychology; Introspection

Further Readings

**SEMANTIC DEMENTIA**

Semantic dementia is a brain disorder characterized by progressive loss of world knowledge and conceptual understanding, which results from degeneration of the temporal lobes of the brain. The disorder typically affects people in the 6th to 8th decade of life. It is part of a spectrum of focal dementia syndromes affecting the frontotemporal lobes and is pathologically distinct from the more common degenerative dementia, Alzheimer’s disease. This entry describes the clinical symptoms of semantic dementia and discusses factors that influence semantic loss, the capacity for relearning, and people’s awareness of their semantic impairment.

The earliest symptoms are typically in the realm of language. People have difficulty remembering the names of things and understanding words. In conversation, they may make semantic errors, such as referring to a sheep as a dog, and use words overinclusively (e.g., *water* to refer to a wide range of liquids). Incorrect word usage reflects a loss of conceptual discrimination between related terms. On hearing some words (e.g., *sheep*), people with semantic dementia may ask what they mean, indicating that the problem is not simply one of word retrieval but a progressive loss in the person’s knowledge of vocabulary. Ultimately, only a few stereotyped words or phrases remain. Nevertheless, people speak fluently and effortlessly within the confines of their increasingly restricted vocabulary. There is no effortful word search because words are no longer available to be sought.

The conceptual loss is not limited to words but encompasses the person’s fund of knowledge relating to all sensory modalities. It affects the ability to recognize objects, faces, nonverbal environmental sounds, tactile, olfactory, and gustatory stimuli. Thus, people may no longer recognize fruits and vegetables in the supermarket, understand the significance of the sound of rain on the windowpane, or recognize the smell of coffee or taste of a lemon. These difficulties do not reflect a problem in sensory perception: People perceive and discriminate sensory stimuli entirely normally. The problem is in ascribing meaning to those percepts. They have lost their semantic associations. In contrast to the profound breakdown in semantic memory, day-to-day autobiographical memory is relatively well preserved, providing a striking contrast to the picture in classical amnesia.

What Is Lost and Retained?

Semantic loss is not all-or-none. During the course of the illness the person will know some things and not others and may have partial, degraded knowledge of a concept. Understanding the factors affecting what is lost and retained potentially informs understanding of the cerebral representation of semantic knowledge.

Modality Effects

If conceptual loss is multimodal, does this mean that information relating to different modalities degrades in parallel? The evidence is controversial. It is common for a person still to recognize the meaning of an object (e.g., a cup) while failing to recognize its verbal label (*cup*). This might be taken as evidence for dissociations between knowledge in different modalities, yet it also might simply reflect differences in task difficulty. Visual stimuli provide clues to meaning that are not available in the word.
A picture of a cup suggests a form of container that can be handled; a picture of a dog suggests an animal by virtue of the presence of legs, ears, and tail. Nevertheless, dissociations have also been detected in the recognition of people’s faces and names; people with more left temporal lobe atrophy have more difficulty recognizing names than faces, and those with more right temporal lobe degeneration have the reverse. Name-face dissociations are less easy to explain in terms of inherent difficulty, a phenomenon that has been interpreted as evidence for different contributions to semantic knowledge from the two cerebral hemispheres.

Frequency, Familiarity, Typicality

The frequency of a word, the familiarity of objects, and the typicality of features in defining a category are strong predictors of performance. Thus, people with semantic dementia are more likely to recognize the common word dog than the uncommon word antelope. They are more likely to recognize the typical feature legs as being an attribute of an animal than the atypical feature hump. Nevertheless, generic measures of frequency, familiarity, and typicality are not sufficient to explain what a person knows. There is also a strong effect of personal familiarity. People show better retention of concepts (words, objects, ideas) relevant to their daily life experience than those that have no personal relevance. This finding suggests a much closer relationship between a person’s semantic knowledge of the world and their autobiographical (episodic) memory than is often acknowledged. What the person experiences has a strong influence on what that person knows.

Learning in Semantic Dementia

People with semantic dementia are able to learn. They can reacquire lost words and relearn the function of objects. However, that knowledge is tenuous and depends on constant rehearsal and applicability to the person’s daily life. Moreover, reacquired knowledge (e.g., that a particular animal is a dog) does not mean that the person’s concept is normal. There is poor generalization to other instances and contexts (e.g., other dogs).

Awareness and Insight

People with semantic dementia are aware that there is something wrong and typically complain that they “cannot remember things.” Nevertheless, they tend to underplay difficulties and typically do not show the great frustration and distress sometimes seen in aphasic patients with severe word finding difficulties. People with semantic dementia are unable to appreciate the magnitude of what has been lost because they no longer have available their prior world knowledge as a comparator.

Conclusion

People with semantic dementia inhabit a shrinking conceptual world. Their understanding becomes progressively narrowed and personalized, limited to the particular instances encountered in their daily life. Semantic dementia provides important insights into how concepts are represented and the relationship between knowledge and experience.

Julie Snowden

See also: Aphasia; Memory, Neural Basis; Semantic Memory; Semantic Memory, Computational Perspectives

Further Readings


SEMANTIC MEMORY

Semantic memory refers to the reservoir of concepts and propositions that you know and that are not tied to any particular time or place. So, for instance, you know that $2 + 2 = 4$, or that carrots are orange, or that there are 50 states in the United States, but you probably don’t remember the specific occasion
on which you learned those facts. Similarly, there are things that you probably don’t know like the 43rd element of the periodic table or the air speed velocity of an unladen European swallow—and when you don’t know something you can often determine that you don’t know it very quickly.

Semantic knowledge can be divided into two main kinds: concepts and propositions. A concept is a mental representation of something—like a dog, or a book, or running, or the color red. Propositions link concepts in statements that are either true or false, such as “The dog was running,” or “The book is red.”

During the 1960s and 1970s, a series of models were developed to explain how semantic memory is organized. In this entry, the three most influential of these models—the hierarchical model, the feature overlap model, and the spreading activation model—will be described.

The Hierarchical Model

According to the hierarchical model, concepts are organized into tree structures. So the class of animals is divided into birds, fish, and so forth. The bird concept is divided into canary, ostrich, and so forth, and the fish concept into shark, salmon, and so forth. Propositions defining the characteristics of a concept are stored with those concepts in such a way as to minimize duplication (a principle called cognitive economy). For instance, the proposition “can sing” would be stored at the canary node in the tree structure but not at the ostrich node. The proposition “can fly,” however, would be stored at the bird concept, because most birds can fly, so by storing it further up the tree, one could avoid duplicating the “can fly” proposition for all the individual birds. Of course, then there is a problem for the ostrich, which cannot fly, and so the proposition “cannot fly” must be stored with the ostrich. As there are very few birds that cannot fly, it is still more efficient to store these exceptions separately than to duplicate the “can fly” proposition for all birds.

To access a memory, the model proposes that people identify the critical concept (e.g., canary) and then traverse the tree to find the concept or proposition that they need. So, for instance, to verify that “a canary can fly,” one would first locate the canary node, note that no flying proposition is connected to that concept, traverse up the tree to the bird node and note that the “can fly” proposition is connected to the bird node, and then make a response. Researchers found that regardless of where the concept was in the tree it took about the same amount of time to go up a level and about the same amount of time to access a proposition—compelling evidence for the hierarchical structure of semantic memory.

However, there is a critical problem. For false statements that are explicitly encoded in semantic memory—that ostriches don’t fly, for instance—one can access the ostrich node, find the relevant proposition, and make a rapid response. However, many false statements are unlikely to be encoded explicitly in semantic memory. For instance, it is unlikely that you have ever had to think about the fact that copper is not an animal, so you would have no opportunity to encode that fact. The only way to deduce that statements like this are false is to search the tree and discover that there is no animal concept above the copper concept. That implies that verifying false statements should, in general, be very slow, but this is not the case.

The Feature Overlap Model

To address this and other problems, the feature overlap model was proposed. According to this account, concepts consist of collections of features. So the concept of a robin might consist of the features, *is biped, has wings, has red breast, and not domesticated*. When trying to verify statements like “a robin is a bird,” the features of the concept robin are compared to the features of the concept bird. Because there is a lot of overlap, one can quickly conclude that a robin is a bird. “A robin is an animal” is verified a little more slowly because there are fewer features in common. Furthermore, if there are very few features in common, people can make a fast no response. So, the statement “copper is an animal” is rejected quickly because copper and animal do not share many features, whereas “a tree is an animal” is rejected a little more slowly because trees and animals are both living things.

The feature overlap model proposes that features come in two types—defining features that all instances of a concept have and characteristic features typical of the concept but which are not present in all cases. However, generating defining features can be problematic. Try to name one feature that
all games have and you will see the difficulty—there always seems to be an exception.

### The Spreading Activation Model

The spreading activation model was designed to address the weaknesses of the earlier models. It is similar to the hierarchical model in that it consists of concepts connected together. However, in the spreading activation model these concepts are not organized into a tree but rather any two concepts that are related can be connected. Furthermore, rather than traversing the tree to find relevant information, in the spreading activation model, activation spreads through all connections simultaneously. If two concepts are strongly related more activation flows between them and if a concept has many connections emanating from it then the activation is divided between these paths—as if activation were a liquid flowing through pipes of different sizes connected in a web.

The spreading activation model was able to account for fast false judgments and did not require any distinction between defining and characteristic features. Although thinking about semantic memory today tends to focus on new computational algorithms, the spreading activation model still plays a significant role in our understanding of the semantic memory system.

In all the models discussed in this entry, it is left up to the theorist to determine the content of the representations employed. Starting in the late 1980s, attention turned toward automatically constructing representations from large text corpora using computational methods. Today, these models play a role not only in theorizing but in real world applications such as Internet searches and automatic essay grading.

*Simon Dennis*

**See also** Categorization, Neural Basis; Categorization, Psychological Perspectives; Classical Theory of Concepts; Concepts, Development of; Representations, Development of; Semantic Memory, Computational Perspectives

**Further Readings**


**Semantic Memory, Computational Perspectives**

Semantic memory refers to factual or conceptual knowledge that is not related to any given personal episode. For instance, the fact that Ottawa is the capital of Canada is a piece of knowledge that you may retain, without necessarily being able to identify when and where you learned it. Similarly, you may know that a poodle and a terrier are similar to each other without ever having been told that they are but rather by virtue of the fact that they fit together in a conceptual representation of the world because they are both small dogs.

The examples given above illustrate the two main kinds of semantic information—conceptual and propositional. A concept is a mental representation of something. So one might have the concept of a dog that becomes active when one sees a poodle, smells a Labrador, is licked by a terrier, or talks about greyhounds. Concepts can also include actions such as running and properties such as red or quickly. Propositions join concepts together into units of mental representation that are capable of having a truth-value. So the concept of a dog is neither true nor false. But the proposition that *dogs have legs* is typically true in our world, although one could imagine a world of legless dogs in which it was not true. Similarly, Ottawa is the capital of Canada, but it is easy to imagine a world in which Toronto was instead.

Early models of semantic memory relied on representations of conceptual and propositional knowledge that were supplied by the theorists. Starting in the late 1980s, however, attention turned to how knowledge could be extracted automatically from exposure to a corpus representing human experience. We will focus on these models starting with those that extract conceptual knowledge and then considering those that extract propositional knowledge.
Conceptual Knowledge

The earliest and most prominent of the models that extract conceptual knowledge is latent semantic analysis (LSA). Introduced by Scott Deerwester, Susan Dumais, George Furnas, Thomas Landauer, and Richard Harshman, LSA derives meaning using statistical computations applied to a large corpus of text. Semantically similar words tend to appear in similar documents. By observing which words appear in which documents in the corpus, LSA defines a set of mutual constraints. These constraints can be solved using singular value decomposition—producing vector representations of both words and documents. The similarity of these vectors is then used to predict semantic similarity.

LSA has been shown to reflect human knowledge in a variety of ways. For example, LSA measures correlate highly with humans’ scores on vocabulary tests, mimic human category judgments, predict how rapidly people are able to access words, and estimate passage coherence. In applied domains, LSA has been used to aid information retrieval, guide discussion forums, provide feedback to pilots on landing technique, diagnose mental disorders, select candidates for jobs, and allow automated tutors to understand the input they receive from students. The most surprising and controversial application of LSA has been its use in automated essay grading. Using the semantic vectors provided by LSA, it is possible to compare novel student essays to essays that have already been graded. If the new essay is most similar to the A essays it is awarded an A, and so forth. The accuracy of LSA at this task is remarkable. It has been consistently shown to correlate with human markers at rates equivalent to the agreement between humans.

While LSA is the best studied of statistical semantics models, a number of other alternatives exist. These include the vector space model, hyperspace analogue to language (HAL) model, the topics model, and sparse nonnegative matrix factorization. Each of these models has its strengths and weaknesses. The topics model has been shown to provide a good account of human free association norms, HAL does a good job of modeling deep dyslexia, while the vector space model, sparse nonnegative matrix factorization, and LSA perform best on document-similarity rating tasks. The topics model and sparse nonnegative matrix factorization have the advantage that they produce vectors with dimensions that tend to be easily interpretable. As of yet, however, no single system addresses all the phenomena of conceptual semantic memory.

Networks of Concepts

Concepts do not exist in isolation but rather are connected together in a web of relationships. Mark Steyvers and Joshua Tenenbaum showed that semantic networks (e.g., the network of associates that people generate) have a small-world structure characterized by sparse connectivity, short average path lengths between words, and strong local clustering. Furthermore, most words have relatively few connections, while a few have high connectivity. These observations are important because they provide constraints on the nature of the process by which conceptual structures grow. A simple model in which new concepts are preferentially attached to already well-connected concepts—a rich-get-richer approach—is sufficient to generate the kinds of graphs observed in adult data.

Propositional Knowledge

What all automated models of conceptual structure have in common is that they assume that documents are bags of words. That is, they do not take into account the order with which words appear and consequently cannot address the question of how propositional knowledge is extracted. Clearly, the sentence “John loves Mary” is not equivalent to the sentence “Mary loves John.” In the first case, John is the lover and, in the second, Mary is. To capture propositional knowledge, one must take into account word order, particularly in languages like English in which word order plays a pivotal role in determining who did what to whom.

Constructing a proposition involves assigning semantic roles (like “lover”) to the entities described in a sentence (like “John”). The problem has been studied at least since Panini, an Indian scholar who worked around 500 BCE. Today, the majority of effort in this area assumes that one has an existing set of semantic roles and access to a corpus in which the roles have been identified before attempting to label new sentences. What this does not explain, however, is how people are able to induce the set of semantic roles in the first instance, or how they
are able to learn when they are not provided with labeled training data.

Simon Dennis provided an answer to this question that revolves around the distinction between syntagmatic and paradigmatic associations. Syntagmatic associations occur between words that appear together in utterances (e.g., run fast). Paradigmatic associations occur between words that appear in similar contexts but not necessarily in the same utterances (e.g., deep and shallow). To understand the logic of the model, suppose we have the following corpus:

- John loves Mary
- Todd loves Sarah
- Barack loves Michelle
- Who loves Sarah? Todd
- Who loves Michelle? Barack

and then present it with the question “Who loves Mary? xxx.” The job of the model is to fill in the “xxx” slot indicating that it understands that John is the lover. To begin with, the model uses syntagmatic associations to determine that words like Todd and Barack could appear in the xxx slot. This pattern—{Barack, Todd}—represents the lover role. Note a similar pattern occurs in the John slot when the model is processing the sentence “John loves Mary.” As a consequence, paradigmatic associations form between the {Barack, Todd} pattern and John. The paradigmatic mechanism in itself, however, would not suffice to predict John, as Barack and Todd are also associated with the lover pattern. Only John has an associative connection to Mary, however, and the additional support afforded by this connection favors John.

The syntagmatic paradigmatic mechanism has been shown to be capable of answering simple questions about tennis matches. Taking naturally occurring text from the Association of Tennis Professionals website, the model was able to complete questions of the form. “Who won the match between Sampras and Agassi? xxx.” Particularly interesting was the fact that the model takes advantage of the systematic occurrence of words through the corpus that occurs as a consequence of the causal relationships between events. As a consequence, it can determine results even when they are not explicitly stated. For instance, the model answered the question “Who won the match between Kiefer and Safin?” based on the sentence “Safin, Kafelnikov surge toward hometown showdown,” without any explicit knowledge that players can only surge toward a showdown when they win. Rather, it relies on the fact that the pattern of players that surge toward hometown showdowns overlaps with the pattern of players that win matches. This kind of inference may be important in understanding how people acquire the impressive stock of commonsense knowledge necessary for comprehending discourse.

Efforts to create computational models of how propositional information is extracted from language are in their infancy. What is clear, however, is that there is a great deal of information that can be extracted to create large-scale models of semantic memory. Advances in computing power coupled with the development of more sophisticated statistical models are providing a deeper understanding of conceptual and propositional structure and how it is acquired.

Simon Dennis

See also Category Learning, Computational Perspectives; Semantic Memory

Further Readings


Sequential Memory, Computational Perspectives

Sequential memory, also called order memory, refers to the process of remembering a sequence of letters, words, digits, pictures, or sounds (here referred to
as items) and reproducing these items in the same order. Sequential memory differs from item memory in that the latter is involved in remembering and reproducing a list of items in any order. Sequential memory is involved in tasks such as reporting back a telephone number and learning to spell a new word. Computational models of sequential memory tackle two questions: How are sequences stored in memory, and how are the items of a sequence selected for output?

**Storing Sequences**

The earliest view about how the mind stores arbitrary sequences of items is referred to as chaining. With chaining, each item is connected with a forward link to the next item. When the sequence has to be retrieved, only the very first item needs to be retrieved and then by following the successive links, the whole sequence is reproduced.

Error patterns seen in data challenge the chaining hypothesis. When performing a task of sequential memory, errors are made that provide insights into how sequences are stored. Imagine that M V F Q S D is the sequence of letters that needs to be remembered by a person. Omissions are errors when the person does not report all the items: M V F _ S D. Transpositions are items reported in the wrong position: M V F S Q D. Intras are items that were not presented in the original sequence: M V F P S D. These examples illustrate a problem with the chaining hypothesis, the strong version of which assumes that the cue for the next item to be reported is the just-reported item. With an omission and intrusion, the relevant cue (in the example, Q) is absent and the person should not be able to report the subsequent items (continuing with S), but experiments show that people do continue. As in chaining models, the Q is linked to S, which is linked to D; after incorrectly reporting the S, the model would continue forward with the D and not backward with the Q. However, experiments show that transpositions (reporting Q after incorrectly reporting S) are the most common errors in serial recall.

Computational models that are able to capture the pattern of transposition errors are compound chaining, positional, and ordinal models. Compound chaining models are models that assume that storage of a sequence involves adding the items to an amalgamation of previous items, which then gets chained to the next item. Positional models are models in which the items are encoded in a gradually changing context representation. This gradual change leads to nearby items being encoded in similar context representations. During retrieval, swaps between nearby items are then possible. Ordinal models are models that assume a gradient over all the items, with the first item being encoded more strongly than the second item, which is encoded more strongly than the third, and so on. During retrieval of items, transpositions will occur for nearby items, as they have similar levels of memory strength.

Two types of errors provide support for the view that the mind stores items relative to (sub)sequence boundaries. When the above sequence is split in two between the third and fourth item, transpositions occur between items that occupy the same intragroup position: M V D—Q S F. This error is called an interposition. Similar errors occur across sequences. When reporting the current sequence, an intrusion can occur with an item that was studied in the previous sequence. The intruding item tends to have been studied in that same position in the previous sequence. This error is called a protrusion. Computational models can produce these types of errors by resetting the context representation at the beginning of each (sub)sequence.

**Reproducing Sequences**

One way of producing a sequence is by retrieving from memory one of the items and, if it is the first item, reporting it. After this, the next item is retrieved and output, using a direct (as assumed by chaining models) or indirect link (via the changing context representation) between the first and second item, and so on. The current dominant view is that producing sequences follows a two-stage process. For example, in typing, the time between two consecutive key presses is much shorter than would be expected by a scenario in which the next key press starts when the previous one has completed. Instead, the findings are consistent with a scenario in which several fingers are moving toward the relevant keys, and only one finger is allowed to complete the response. The simultaneous activation of responses (the first stage), together with the selection of one response (the second stage), leads to smooth production of the items (key presses); this two-stage process has been called competitive queuing.
In computational models of sequential memory, competitive queuing is central to sequence production. The dynamics are such that after cues activate the target items, the item with the strongest support is selected and output. After output, this item is inhibited so that the next strongest item can be selected and output. As ordinal models have a rank ordering of items from strong to weak, these models produce the items in the forward order. As positional models do not have a gradient that rank orders the items, the context signal is replayed and candidate items activated. The strongest item is the most likely item that was encoded in that position and will be output (unless it was already output erroneously).

Eddy J. Davelaar

See also Natural Action Selection, Modeling; Serial Order Memory, Computational Perspectives

Further Readings


Serial Order Memory, Computational Perspectives

Serial recall requires people to accurately remember and recall the order of sequences of information such as letters, digits, and spatial locations over short periods of time. This task is of interest as one of a number of tasks used by both researchers and clinicians to tap short-term memory abilities; when combined with an interleaved processing task (e.g., a choice reaction time task), the combination is a good predictor of higher level cognitive abilities such as reasoning. Serial recall is especially popular given the specific focus on memory for the order of information and has been argued to be fundamentally involved in learning the phonology of new words. Reflecting this interest, a number of models of serial recall have been developed to account for serial recall performance. The theoretical development in this area has been impressive, with models accounting for a comprehensive set of data at a fine level of detail. The key phenomena accounted for by these models include primacy and recency, whereby memory accuracy declines across positions in the sequence with the exception that the last one or two items in a sequence are better remembered; the locality effect, whereby an item recalled in an incorrect position will nonetheless tend to be recalled in a nearby position; and the phonological similarity effect, where verbal materials that rhyme or share a number of phonemes will be less well recalled, particularly because of worse memory for the ordering of those materials in the sequence.

Representing Order

One basic issue addressed by these models is how the order of elements in a sequence is represented. Figure 1 shows three general schemes of representation of order in contemporary models of serial order. In the top scheme, chaining, adjacent elements of a sequence are associated in memory. Once an item is recalled, the following item can be accessed by using the recalled item to cue the next item via the pair association. Although Stephan Lewandowsky and Bennet Murdock showed that the chaining model could account for many of the key phenomena described above, later work has challenged this model. A particularly troublesome finding is that alternating rhyming and nonrhyming items in a sequence does not harm—and may even enhance—recall of the dissimilar items. Chaining models predict that confusions of rhyming items should be followed by confusions of nonrhyming items: When a rhyming item is recalled in the wrong position, the following nonrhyming item should move with it and thus be incorrectly ordered.

This led to the development of primacy gradient models, in which order is represented as a gradient of activation or encoding strength across items. These models successfully account for many of the key phenomena and can allow nonrhyming items to be protected from confusions between rhyming items by assuming that phonological confusions
Serial Order Memory, Computational Perspectives

occur in a stage downstream from the primacy gradient. One challenge to these models is the effect of grouping on serial recall: Temporally grouping a sequence by placing pauses between subsequences produces primacy and recency within subsequences and also leads to specific patterns of confusions in between groups, the most telling of these being that elements that appear at the same position in different subsequences are more likely to be confused. These grouping effects imply some form of factorial or hierarchical representation in short-term memory, taking these data beyond the limits of the unidimensional, strength-based, primacy-gradient models.

**Positional tagging** models address this shortcoming by explicitly introducing a multidimensional representation of order that is located separately from the item representations. Incoming sequence elements are associated to successive positional tags with proximate tags overlapping to a larger extent than tags separated by a large number of intervening tags. Grouped lists are represented jointly by tags coding the position of an element in a group and some coarser representation of that element in the entire list. One debate is whether these tags code the real-time occurrence of items, whether they instead represent the relative timing of items, or whether these tags are purely ordinal and driven by successive encoding events.

**Mechanisms of Encoding and Retrieval**

Aspects of serial recall data imply additional mechanisms at play in the encoding and retrieval of sequences. To explain the generation of the primacy gradient, Lewandowsky and Simon Farrell suggested that the strength of encoding of elements is determined by the novelty of each element with respect to those already in memory. Successive elements will be less and less novel, thereby leading to a primacy gradient. Additionally, most models assume some form of response suppression whereby those elements that have already been recalled are reduced in activation or competition to prevent their further recall. This is used to explain the unwillingness of participants to repeat elements in their output even when sequences do actually contain repetitions.

Contemporary models also address the role of serial recall in the long-term learning of sequences. Recent work aims at accounting for the Hebb effect—the improvement in performance following from (nonimmediately) repeated presentations of a list—and relating this to the claimed purpose of verbal short-term memory as a mechanism to learn new words (sequences of phonemes). The model of Neil Burgess and Graham Hitch assumes that the associations between position tags and elements are retained into following sequences and that an incoming sequence is matched to all those previously learned sequences. The Hebb effect results from the fact that repeating a sequence will lead to the continued reuse and strengthening of the specific tag-element associations for that sequence.

*Simon Farrell*

**See also** Sequential Memory, Computational Perspectives; Working Memory

**Further Readings**


**Similarity**

An enormous amount of ink has been spilled in the psychology literature on the topic of similarity. There are two reasons that this seemingly intuitive and prosaic concept has been the subject of such intense scrutiny. First, there is virtually no area of cognitive processing in which similarity does not seem to play a role. William James observed in 1890 that “this sense of Sameness is the very keel and backbone of our thinking” (p. 459). Ivan Pavlov first noted that dogs would generalize their learned salivation response to new sounds as a function of their similarity to the original tone, and this pattern of generalization appears to be ubiquitous across species and stimuli. People group things together based on their similarity, both during visual processing and categorization. Research suggests that memories are retrieved when they involve similar features or similar processing to a current situation. Problem solutions are likely to be retrieved from similar prior problems, inductive inference is largely based on the similarity between the known and unknown cases, and the list goes on and on.

Psychology clearly has an enormous stake in similarity. In fact, the depth of this stake makes it that much more unsettling that similarity can be such a slippery and temperamental construct. Perceived similarity can vary considerably with context. In fact, the act of comparison itself can change people’s representations, leading them to construe things as more comparable by reinterpreting the given features or even creating new features. Similarity ratings are often asymmetric; for instance, people view North Korea to be more similar to China than the reverse. Likewise, similarity and difference are not always the inverse of one another. When given a choice between XX and OX, people choose XX as both more similar to and more different from OO. Similarity can change systematically with temporal distance and physical distance, and there is a growing body of evidence for consistent individual differences in which kinds of features drive a person’s similarity judgments. Judgments of similarity can increase with simple association. For example, coffee is judged to be similar to cream because it is contextually associated with cream. In the absence of objective ways to measure psychological similarity, researchers are left to rely on participants’ subjective judgments of their own processing, which are potentially unreliable, or on data from tasks in which similarity is proposed to play a role, leading to conclusions that are potentially circular. Thus, it is possible that similarity is both the most essential and the most problematic construct in cognitive science. This entry discusses some of the important findings from the extensive literature on similarity and the ways in which psychologists have attempted to address a topic so critical to our understanding of the mind and yet so elusive.

**Theories of Similarity**

As one might expect given the importance of similarity for thinking, it is understandable that there have been several attempts to formalize the process of determining similarity. These formal accounts stipulate how similarity is to be empirically measured and provide theoretic accounts of how similarity should be conceptualized. The resulting models have had a large practical impact in knowledge representation, automatic pattern recognition by machines, search engines, data mining, and marketing (e.g., online stores that recommend new products to you based on your similarity to previous buyers). The entry surveys four of the most prominent models of similarity: geometric, feature based, alignment based, and transformational.

**Geometric Models and Multidimensional Scaling**

Geometric models work under the premise that what it means for two things to be similar is for them to be close to one another in a psychological space. These approaches are exemplified by the statistical modeling technique of multidimensional scaling (MDS). MDS models represent similarity relations between entities as a geometric model that consists of a set of points embedded in a space. The input to MDS routines may be similarity judgments, confusions
between entities, patterns of co-occurrence in large samples of text, or any other measure of pairwise proximity. Most straightforwardly, participants may be asked to judge how similar every object in a set is to every other object. The output of an MDS routine is a geometric model of the set of objects with each object represented as a point in an n-dimensional space. The similarity between a pair of objects is inversely related to the distance between the objects’ points in the space. In MDS, the distance between points i and j is typically computed by

\[ \text{dissimilarity}(i,j) = \left[ \sum_{k=1}^{n} |X_{ik} - X_{jk}|^r \right]. \]

where \( n \) is the number of dimensions, \( X_{ik} \) is the value of dimension k for item i, and r is a parameter that allows different spatial metrics to be used. With \( r = 2 \), as in standard Euclidean geometry, the distance between two points is the length of the straight line connecting them. If \( r = 1 \), then distance involves a city-block metric in which the distance between two points is the sum of their distances on each dimension; “short-cut” diagonal paths are not allowed. Empirically, the Euclidean distance measure typically fits human data better when the stimuli being compared consist of perceptual dimensions that are psychologically fused together. For example, brightness is a subjective dimension related to the amount of luminance energy coming off of an object. Saturation is a subjective dimension related to the amount of monochromatic light mixed into a color. Brightness and saturation are psychologically fused in the sense that it is difficult to pay attention to brightness differences between objects without also being influenced by saturation differences. For objects differing on saturation and brightness, their subjective similarity is best measured by a distance calculation that fully integrates saturation and brightness differences together—namely, \( r = 1 \). Conversely, if objects differ on brightness and size, then their similarity is best measured by computing their distance on brightness and then adding this to their distance on size—namely, \( r = 2 \).

A classic example of MDS comes from Ed Smith, Edward Shoben, and Lance Rips’s study of animal concepts. They asked participants to provide similarity ratings on many pairs of birds or other animals. Submitting these pair-wise similarity ratings to MDS analysis, they obtained the results shown in Figure 1. The MDS algorithm produced this geometric representation by positioning the birds in a two-dimensional space such that birds rated as being highly similar are very close to each other.

One practical limitation of MDS is that obtaining all pairwise similarity ratings among a large set of objects requires a substantial commitment of time and effort. If similarity ratings are used as the input to MDS, then standard \( N^2 \) ratings are required for \( N \) objects. This number is halved if one assumes that the similarity of A to B is the same as the similarity of B to A. Even with this halving, the number of ratings still becomes prohibitively large as \( N \) increases. Fortunately, automated techniques for analyzing large corpora of text can provide input to MDS models instead of relying on “manually” provided ratings. Using this method, latent semantic analysis is a computational approach to word meaning that has received considerable recent attention. It bases word meanings solely on the patterns of co-occurrence between a large number of words in an extremely large text corpus such as an encyclopedia or thousands of email messages. It employs the mathematical analysis tool of singular value decomposition (SVD) to create vector encodings of words that efficiently capture their co-occurrences; these encodings represent each word by an ordered set of numbers—that is, a vector. The similarities between two words’ vectors efficiently capture their co-occurrences. If two words, such as cocoon and butterfly, frequently co-occur or enter into similar patterns of co-occurrence with other words, their vector representations will be highly similar. The meaning of a word, its vector in a high dimensional space, is completely based on the contextual similarities among words. Within this high dimensional space, Thomas Landauer and Susan Dumais conceive of similarity as the cosine of the angle between two words rather than as their distance. By using these new techniques, it is possible to create geometric spaces with tens of thousands of words.

**Featural Models**

In the 1970s, it was observed that subjective assessments of similarity do not always satisfy the assumptions of geometric models of similarity:

- **Minimality:** \( D(A,B) \geq D(A,A) = 0 \)
- **Symmetry:** \( D(A,B) = D(B,A) \)
- **The Triangle Inequality:** \( D(A,B) + D(B,C) \geq D(A,C) \)
where $D(A,B)$ is interpreted as the dissimilarity between items A and B. Minimality captures the simple ideas that no object should be more dissimilar to itself than it is to another object and that the dissimilarity of an object to itself should be 0. Symmetry captures the notion that the dissimilarity of Object A to B should be the same as the dissimilarity of Object B to A. Whatever Lady Gaga’s similarity is to Madonna should be the same as Madonna’s similarity to Lady Gaga. The notion behind the triangle inequality is that the length of the direct path from A to C should be no longer than the path from A to B plus the path from B to C. For similarities, this means that if a red square (A) is fairly distant from (dissimilar from) a blue circle (C) then the red square’s distance to a red circle (B) and the red circle’s distance to the blue circle cannot both be very short, otherwise the two-legged detour route from A to C going through B will be shorter than the direct route from A to C.

In fact, violations of all three assumptions have been empirically obtained. In response to these violations of the geometric model of similarity, some researchers have proposed fixes that allow, for example, the local density of objects in a region to warp the calculation of distance/dissimilarity. More radically, in 1977 Amos Tversky proposed to model similarity according to matching and mismatching features rather than distance on psychological dimensions. In his model, entities are represented as a collection of features and similarity is computed by

$$S(A,B) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A)$$

where $S(A,B)$ is the similarity of Object A to Object B and is expressed as a linear combination of the measure of the common and distinctive features. The term $(A \cap B)$ represents the features that Objects A and B have in common. $(A - B)$ represents the features that A has but B does not. $(B - A)$ represents the features of B that are not in A. And the $\theta$, $\alpha$, and $\beta$ are weights for the common and distinctive components, reflecting how important each component is for determining similarity. For example, in Figure 2 we imagine comparing robots (A) to zombies (B). This would be accomplished by first determining all the features of each of these two objects. Then, their similarity is calculated to be a positive function of

![Figure 1](source:image_url)

**Figure 1** Two multidimensional scaling (MDS) solutions for sets of birds (A) and animals (B)


*Note: The distances between the animals in the space reflect their psychological dissimilarity. Once an MDS solution has been made, psychological interpretations for the dimensions may be possible. In these solutions, the horizontal and vertical dimensions may represent size and domesticity, respectively.*
their shared features and a negative function of the features possessed by robots but not zombies \((A - B)\) and of the features possessed by zombies but not robots \((B - A)\).

Alignment-Based Models

MDS and featural models make different assumptions about similarity, but they also share a number of similarities. An important commonality between geometric and featural representations is that both use relatively unstructured representations with entities structured as sets of unrelated features or dimensions. However, entities such as objects with parts, real-world scenes, words, sentences, stories, scientific theories, and faces are not simply a “grab bag” of attributes. A dog biting a man is not the same thing as a man biting a dog, even though they both feature a dog, a man, and biting. How these elements are related matters. Partly in response to the problems that geometric and featural models have with structured descriptions, a number of researchers have developed alignment-based models of similarity. In these models, comparison does not just involve matching features but also involves determining how elements correspond to, or align with, one another. Matching features are aligned to the extent that they play similar roles within their entities. For example, a man wearing a black tie and a woman with black shoes both share the feature *black*, but this matching feature may not increase their similarity much because the man’s tie does not correspond to the woman’s shoes. Drawing inspiration from a structure-matching model of analogical reasoning by Dedre Gentner, in alignment-based models, matching features influence similarity more if they belong to parts that are in correspondence, and conversely, parts tend to be placed in correspondence if they have many features in common.

Transformational Models

A final approach to similarity maintains that the similarity of two objects is directly related to the number of transformations required to turn one object into the other. An important step for these models is to specify what transformations are possible.

Researchers in artificial intelligence have claimed that objects are recognized by being aligned with memorized pictorial descriptions. An unknown object will be placed in the category that contains the candidate model with which it best aligns. The alignment operations rotate, scale, translate, and topographically warp object descriptions.

According to Ulrike Hahn, Nick Chater, and Lucy Richardson, the similarity between two entities is based on how complex the sequence of transformations is that changes one entity to the other. The simpler the transformation sequence, the more similar the entities are assumed to be. For example, the transformational complexity connecting \(1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8\) and \(2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 8 \ 9\) is small because the simple instruction “add 1 to each digit” suffices to transform one into the other. Experiments demonstrate that once reasonable vocabularies of transformation are postulated, transformational complexity does indeed predict subjective similarity ratings. Furthermore, when a new transformation is learned that turns Object A into Object B, A is seen as more similar to B.

Conclusions

It might be argued that all four of the above approaches err on the side of treating similarity as a unitary phenomenon. It could well turn out that calculating similarity is fundamentally different for
different kinds of entities. Taken to an extreme, this notion raises the possibility that similarity is not a coherent notion at all. Like the terms bug or family values, similarity may not pick out a consolidated or principled set of things. Consistent with this possibility, it may be that much of the real theoretical work in the future will be achieved by determining what counts as the features and relations that underlie similarity assessments for different kinds of entities and in different situations. Nonetheless, one justification for pursuing general theories of similarity is that if they do exist, a large payoff results. Even if similarity is not a monolithic entity, there probably will be common cognitive processes involved in different kinds of comparisons.

Some philosophers have attacked the very notion of similarity as being empty or circular. They have pointed out that the claim that entities A and B are similar is vague and ill defined unless one specifies the aspects under consideration when making the claim. However, part of the power of the notion of similarity is that it integrates over many aspects of entities. All four of the models of similarity can be interpreted as proposals for how the process of integrating across aspects proceeds. By integrating over many aspects, similarity is a powerful tool for cognition because it does not require the cognizer to fully understand exactly what makes entities behave as they do. It only requires that the world is a sufficiently orderly place that similar objects and events tend to behave similarly. This fact of the world is not just a fortunate coincidence. It is because objects are similar that they tend to behave similarly in most respects.

Robert L. Goldstone and Samuel B. Day

See also Analogical Mapping and Reasoning; Categorization, Psychological Perspectives; Category Learning, Computational perspectives; Concepts, Development of; Concepts, Philosophical Issues; Semantic Memory; Semantic Memory, Computational Perspectives

Further Readings


Skill Learning, Enhancement of

Skill learning involves experience that leads to improvements in performance. This concept must be distinguished from the similar concept of knowledge acquisition. For skill learning, performance requires attaining procedures, whereas for knowledge acquisition, it requires attaining facts. However, most, if not all, activities involve both procedural information (skill) and declarative information (facts).

To enhance skill learning, three different aspects of the process should be strengthened: acquisition, retention, and transfer. First, training of the skills should be accomplished quickly and efficiently to reduce costs and save time and effort. Second, the skills should be made as durable as possible so that following training they can still be available at a high level even when they have not been used for a long period of time. Third, the skills should be made as flexible as possible so that they can be applied in different contexts and their use is not restricted to the particular situations encountered during training. Research in experimental psychology has led to
several principles of training that can enhance the efficiency, durability, and flexibility of skill learning.

**Principles of Training**

**Deliberate Practice**

Skill learning usually occurs gradually, with individuals improving their response speed following what is known as the power law of practice. According to this law, practice helps most at the beginning, but even after many trials, practice continues to make individuals faster. There is also a complementary power law of forgetting, which accounts for a gradual decline in performance with the passage of time when there is no opportunity to refresh the acquired skills.

Even though the acquisition and retention of skills are well described by these power laws, learning and memory of skills vary depending on multiple factors. One such factor is type of practice. Not all practice is equal in promoting skill learning. Deliberate practice, which requires a high degree of focus and motivation, is required for achieving expert levels of performance, even among individuals who seem to show a natural talent or aptitude for a particular skill.

**Procedural Reinstatement**

Although acquired knowledge is forgotten very rapidly, learned skills are usually very well retained across periods of disuse. For example, learning a telephone number from a phone book might not survive the 30 seconds it takes to walk from the book to the telephone and dial the number. In contrast, learning how to use a hula hoop might survive years in which no hoop is available so that an individual who learned how to use a hula hoop as a child might be able to use the hoop again perfectly well as an adult, even if there was no intervening practice with the hoop since childhood. However, skill learning is usually not very flexible whereas knowledge acquired in one situation can be easily applied to many other situations. Thus, to continue with the examples, the learned hula hoop skill might not help performance with a hoop that is not the same size as the ones used during childhood, but the knowledge of acquired telephone numbers can be used in many different contexts, including the learning of new sequences of numbers. These observations form the procedural reinstatement principle, according to which declarative information shows poor retention but robust transfer and procedural information shows strong retention but limited transfer.

Evidence supporting this principle comes from experiments showing a high degree of specificity of skill learning. For example, students learned how to use a computer mouse to move a screen cursor from a central location to targets along the screen periphery. The task was made more difficult by changing the relationship between the mouse and the cursor. In one situation the cursor moved left whenever the mouse moved right and vice versa, but the up-down relationship between mouse and cursor movements were intact. Students were able to acquire this skill rapidly and retained it well over a long delay, but in many cases they could not transfer the skill involving one mouse-cursor relationship to that involving another; in fact, sometimes there was interference from learning the first relationship to learning the second. Other experiments show that skills are specific to the context in which they are learned. For instance, students learned how to produce time intervals based on arbitrary units. They were not told how long a given unit was, but they received feedback on their productions, allowing them to become increasingly more accurate. These time intervals were produced in the context of various background tasks, such as the difficult task of repeating the alphabet backward by threes. The difficult tasks lowered performance on time production, but after learning to produce time intervals with one background task, participants could not generalize their skill to producing time intervals with another background task, even when they changed from a difficult background task to no background task at all.

**Variability of Practice**

Especially given the lack of transfer found for skill learning, it is important to consider whether there are training methods that can be employed to help promote skill transfer. One such method that has proven to be effective is based on variability of practice. Training with a variety of tasks usually leads to better transfer performance than training with a single, constant task. This benefit for variable training is sometimes found even when the task used in testing differs from those used during variable practice and is the same as that used during constant practice. For example, practice that varied the target distance at which bean bags were tossed
by children (2 or 4 feet) led to better accuracy at test on an intermediate target distance (3 feet) than did practice that was restricted to the same intermediate target distance. This variability of practice principle has been found to apply to both discrete and continuous motor tasks as well as to various cognitive tasks such as learning new concepts or understanding passages of text.

**Contextual Interference**

Most teachers try to make learning as easy as possible for their students. Indeed students will learn more rapidly under easy conditions than under difficult conditions, and tests of performance immediately after learning will often be better for students who learned under easy conditions than under difficult conditions. However, when testing the same students after a delay or in a new situation, it is often found that performance is better for students who learned under difficult than under easy conditions. In fact, introducing desirable sources of difficulty or interference during training has been shown to be an effective way to promote retention and transfer, following the principle of contextual interference.

The most common way to study contextual interference has been to compare mixed and blocked practice schedules. In blocked practice schedules, practice with a given task or with a given set of materials occurs together in the same block of trials, whereas in mixed practice schedules, practice occurs on each task or with each set of materials in every block. Mixing tasks or materials during training produces interference and, thus, retards initial skill learning but often enhances ultimate skill use.

**Distribution of Practice**

A related way to enhance skill learning concerns the distribution of practice. Practice trials can occur without interruptions or with no time in between trials, in which case they are massed. Alternatively, such trials can occur spaced apart or distributed in time. It has generally been found that spaced practice is better than massed practice, especially with a long retention interval, or delay between the end of learning and the beginning of testing. In fact, optimal performance at test seems to result when the interval between practice trials is approximately equal in length to the retention interval. This conclusion seems to hold even when practice is distributed across sessions rather than across trials within a session.

Another related question concerns the order in which to train tasks that vary in difficulty. According to one line of research, errorless learning should be encouraged. With errorless learning of a motor skill, trainees start with the easiest task and then gradually progress to more difficult tasks. For example, in learning how to putt a golf ball, the trainee might start with the shortest putting distance and then gradually increase the distance. This arrangement of practice trials should be likely to reduce the number of errors made by the trainees relative to other arrangements. It has been argued that when more errors are made during skill learning, performance requires more attention-demanding resources than when errors are minimized, so that distractions and stress cause less performance disruption following errorless learning than following learning that occurs with frequent errors.

**Focus of Attention**

The demands of attention are also relevant to another training principle that specifically involves the focus of attention. It has been shown for a variety of sport skills, again including golf shots, that an external focus of attention leads to better retention than does an internal focus of attention. With an external focus, attention is given to the results of a movement (e.g., where the ball lands) whereas, with an internal focus, attention is given to the movements of the body (e.g., how the arms move during the swing). The claim has been made that an internal focus requires conscious attention to motor movements and such attention impairs automatic mechanisms in the body underlying skilled performance.

**Mental Practice**

Whether using an internal or external focus of attention, practice is necessary for skill improvement. However, there are times when it might be inconvenient or even impossible to practice a skill, especially if the skill requires special equipment not readily available. The question arises as to whether mental practice can provide a reasonable substitute for physical practice in those circumstances. In one study involving data entry, participants practiced typing four-digit numbers on a computer, either by actually typing the numbers or by just imagining their typing movements. For both types of practice, participants increased the speed at which they executed the typing responses, showing
equivalent improvements in typing skill. When practice involved a key configuration different from that at testing, participants who used physical practice but not those who used mental practice suffered from interference. These results suggest that mental practice supports a more abstract representation of the skill than does physical practice.

**Cognitive Antidote**

Although practice usually leads to improvements in skill performance, sometimes performance on a task deteriorates with prolonged work on that task. In fact, for the routine data entry skill, lengthy practice led to both improvement and deterioration, depending on the measure used to assess skill performance. For response speed, performance got better and better with practice, but for accuracy, performance got worse and worse. This pattern illustrates an increasing trade-off of accuracy for speed. The faster speed is easily understood in terms of the principles already discussed, whereas the decline in accuracy can be explained in terms of increasing fatigue, boredom, and task disengagement. The drop in accuracy can be overcome, though, by providing feedback about errors, thus increasing motivation. In addition, accuracy can be enhanced and the decline in accuracy eliminated by requiring an extra cognitive task, such as ending the typing response for each number by hitting either a + key or a − key, with the two keys alternating across trials. This cognitive antidote serves to eliminate the speed-accuracy trade-off otherwise observed as a result of prolonged work.

**Conclusions**

In summary, skill learning can be done efficiently with improved skill performance occurring even following a long delay after practice is completed. However, transferring the skill to new situations is often difficult but can be promoted by following a number of training principles, which recommend employing deliberate practice, increasing the variability of practice, adding sources of contextual interference, using a mixed practice schedule, distributing practice in time, eliminating errors, and using an external focus of attention. Mental practice can substitute for physical practice, and boredom can be overcome by adding a cognitive antidote to routine tasks.

*Alice F. Healy*

**See also** Desirable Difficulties Perspective on Learning; Motor Learning, Practical Aspects; Rehearsal and Memory; Spacing Effect; Spacing Effect, Practical Applications

**Further Readings**


**Sleep and Dreams**

Within a day, all animals and humans show recurring periods of immobility, usually in a characteristic posture and environment, and associated with an apparent loss of responsiveness to environmental
input—they sleep. In contrast to a comatose state, some sensory discrimination is preserved and sleep is reversible when strong stimulation is applied.

Historically, the views that sleep is simply rest and that dreams convey a special meaning has prevailed in most cultures. Dream contents were assigned a mystical dimension—they were seen as messages sent by a god or as representing states unavailable to waking consciousness and foretelling future events, an attitude still popular. Contrary to these subjective interpretations, but instead based on measurable changes of brain activity and information processing mechanisms, research activities in the past decades have clarified different states within sleep, have unraveled many active processes within sleep, and have highlighted neurobiological processes underlying a multitude of sleep-related phenomena in animals and in humans. Until now, several functions of sleep have been proposed, ranging from energy conservation, thermoregulation, and detoxification to brain plasticity processes resulting in functional or structural changes. Current knowledge substantiates that neuronal assemblies are active, reactivated, or even modified during sleep.

This entry reviews the classic and more recently applied methods to study sleep and sleep-related phenomena. These approaches allow us to describe the substantial changes in the brain’s capacities to process external information and brain activation patterns across the different stages of sleep and to investigate the neuronal basis of dream mentation.

Methods

The current understanding of cortical activity across the different stages of waking and sleep mainly derives from surface electroencephalographic (EEG) recordings reflecting cerebral synaptic activity. As a consequence, EEG criteria are generally used to differentiate the increasing depth of sleep on slowing of the EEG rhythms. The formerly held belief of sleep as a cessation of brain activity and annihilation of consciousness was finally overthrown in 1953 when Eugene Aserinsky and Nathaniel Kleitman first described an active brain state recurring in regular intervals within sleep. This state is accompanied by a loss of voluntary muscle control and by rapid eye movements. It is therefore called rapid eye movement (REM) sleep or paradoxical sleep, as it shares many features of neuronal behavior with wakefulness but still represents a sleeping state with high arousal thresholds.

Electrophysiological Sleep Recordings

REM and non-REM (NREM) sleep stages can be differentiated based on recorded EEG, eye movement (electrooculographic [EOG]), and muscle activity (electromyographic [EMG]) criteria. Waking and consciousness rely on activity of neurons in the formatio reticularis (reticular formation) of the brain stem, forming the ascending reticular activating system projecting to the thalamus and cortex. Waking EEG recordings are dominated by fast-frequency activity in the beta (15–30 Hertz [Hz]) and gamma range (30–80 Hz). The electrophysiological signs of NREM sleep—synchronized low-frequency oscillations with high amplitudes in the EEG recordings—seem to confirm the early notion of NREM sleep as a state of rest with cessation of intense cortical activity.

In contrast to NREM sleep, REM sleep shares many signs with wakefulness such as fast-frequency and low-voltage cortical EEG, including gamma oscillations. Whereas slow-wave sleep depends on thalamocortical synchronization, specialized cells in the brain stem called REM sleep-on cells initiate and generate signs of REM sleep. From an evolutionary point of view, REM sleep is exclusively found in birds and mammals—that is, it is unique to endotherm animals with well-developed brains that also express slow-wave activity, with the exception, for example, of dolphins. During ontogenesis (development of the individual), REM sleep is found in relatively high amounts during early development, and a link to brain maturation in phylo- and ontogenesis was proposed. Because of circadian modulation, REM sleep episodes are longest and REM density is strongest at the time of minimal body temperature, usually in the early morning hours. The concurrent inhibition of spinal motoneurons led to the depiction of REM sleep as a “highly activated brain in a paralyzed body,” also lacking proper input processing.

Embedded in circadian and ultradian (occurring more than once in 24 hours) rhythms paced by a circadian clock in the suprachiasmatic nucleus, the sleep stages appear in alternating cycles lasting about 90 minutes in adults, with a predominance of slow-wave sleep in the first and of REM sleep.
in the second half of the night. About 80% of the sleeping time of adult humans is spent in NREM sleep, characterized also by general slowing of body functions. For clinical purposes, sleep recordings are usually extended to polysomnographic recordings with additional measurements of functions such as heart rate, breathing parameters, activity of specific muscle groups, and so forth.

The amount of slow-wave activity as measured by EEG spectral analysis is popular as a measure of NREM sleep intensity, which is sensitive to increased sleep pressure (homeostatic upregulation) after sleep deprivation. Intensity of REM sleep can be quantified in percentage and duration of REM sleep episodes and in REM density as a measure of relative amount of, for example, rapid eye movement activity within REM periods. Quantification of EEG data further encompasses a calculation of coherence between electrode positions as a sign of correlated activity or methods for source localization of the electrical activity recorded at the brain surface.

**Event-Related Potentials**

In addition to these traditional methods of quantifying sleep, mechanisms of information processing can be studied using event-related potentials (ERPs) on stimulation. Here, surface EEG recordings are separated into individual segments and then averaged time-locked to specific stimulus onset times. The averaging process decreases noisy components and highlights stimulus-locked electrophysiological responses in the range of milliseconds to seconds. The components are described by the polarity (positive-negative) and by the latency (amount of time elapsed from the onset) of the local maximum amplitude. Fast components (up to 50 milliseconds [ms]) reflect early signal transduction pathways preceding higher cortical involvement and show longer latencies throughout sleep. Later potentials (e.g., at 300 ms) reflect higher cortical levels of signal processing and display specific alterations during different sleep stages.

**Imaging Methods**

In the recent years, major advances in imaging methods have enabled new insights into brain processes during sleep. Positron emission tomography (PET) using radioactive tracers and the noninvasive methods of functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and high-density EEG recordings provide maps with localization of neuronal activity. These methods highlight the regional specificity of changes in neuronal assemblies linked to sleep-related phenomena.

**Behavioral Assessment and Subjective Reports**

Measuring cognitive and neuropsychological aspects in relation to sleep such as pre- and post-sleep performance allows for a more comprehensive understanding of sleep-related functions. Next to objective measures, information on subjective experiences plays a crucial role, especially when studying dreams, which cannot be directly observed. Data can be collected via questionnaires, structured or semistructured interviews, or diaries. Apart from collecting spontaneous memories of dream contents, some sleep laboratory studies provoke awakenings of the participant from specific sleep stages and immediately obtain information on ongoing mentation. This procedure can decrease forgetting, but issues such as self-censorship or difficulties with verbal description may still be present.

**Information Processing During Sleep**

Missing reactions to environmental changes are a prominent hallmark of sleep. To stay asleep, an organism needs to ignore the disturbing environmental influence. Reactivity decreases during falling asleep and, finally, no reactions can be elicited, which has made sleep appear as an entirely passive state. These changes are paralleled by a fading of consciousness. Thus, the first measurements of sleep depth were based on missing reactions to acoustic stimuli of different loudness. On the contrary, the ability to discriminate stimuli of personal relevance is still preserved, which is very obvious in “mother’s sleep,” when parents awake to even the slightest sounds of their infants. Not only is sensory processing altered during sleep, but sensory information can also alter the stages of sleep, resulting in a reciprocal interaction of sensory input and vigilance.

Information processing in sleep can be studied using single cell studies in animals or by analyzing EEG, evoked potentials, or imaging data in humans. The main areas of research have been altered reactions during sleep and, more recently, changes in memory or problem-solving capacities following sleep.
Mechanisms Underlying Information Processing in Sleep

Studies at the level of single nerve cells have confirmed that during sleep cortical neurons reduce their overall activity compared to wakefulness, switching from continuous, tonic firing to burst-mode pattern (rapid firing followed by silent periods) in NREM sleep as reflected in synchronized EEG delta activity. However, the cerebral cortex is not globally inhibited, and the most striking difference compared to wakefulness is a change to synchronous firing at a low-frequency rhythm. The synchronized burst activity has recently been associated with specific plasticity changes. During the silent periods (hyperpolarized states), less energy is required, similar to an idling motor that needs less energy. It should be noted that many cortical cells fire at high levels also during NREM sleep, confirming active processes within NREM sleep. During REM sleep, cholinergic mediated activation spreads from the pontine structures and activates thalamic, limbic, and cortical areas.

During wakefulness, the thalamus plays a central role in conveying incoming afferent sensory signals—except olfaction—to the corresponding cortical areas, and cortical neurons reciprocally communicate to the thalamus. With increasing depth of sleep, first thalamic and then cortical cells switch to an altered functional state. Prominent graphic elements of NREM sleep EEG in humans with functional significance for information suppression are sleep spindles and K-complexes (KCs). Especially during occurrence of sleep spindles, which are 12 to 15 hertz (Hz) oscillations generated in the thalamic nucleus reticularis, the transfer of sensory signals to the cortex is blocked. However, spindle activity seems to be an autonomous process that even declines with increasing stimulus intensity, in contrast to the cortically generated KCs that can be elicited on all kinds of sensory stimulation.
KCs were first described in 1938 by Alfred Loomis; the name K-complex is derived from “to knock,” as an example of an external acoustic stimulation. KCs are a complex of several inhibitory and excitatory evoked potentials, and have been discussed to reflect sleep-protective mechanisms. Finally, during slow-wave sleep, the slow oscillations go along with a strongly reduced excitability of cortical neurons. Thus, even if sensory information is conveyed to the corresponding cortical area, it cannot be properly processed during the long states of hyperpolarization reflecting neuronal silence.

In contrast to the above described gradual loss of reactivity in NREM sleep, the ability to discriminate complex stimuli is restored during REM sleep. Suppression of information processing is not mediated via thalamocortical inhibition in REM sleep but was suggested to lack reasonable integration because of altered activity in frontal, associative, and limbic cortex. Arousal thresholds appear higher during phasic REM periods when bursts of rapid eye movements appear. External stimuli can be kept in REM sleep mentation up to 15 minutes after presentation.

**Event-Related Potential and Imaging Studies**

By averaging electrophysiological responses, evoked potential studies consistently reveal a decrease of negative and increase of positive cortical potentials as the organism moves from wakefulness to NREM sleep stages, interpreted as reflecting inhibitory action and loss of attentive processes. A very prominent ERP component during wakefulness is the P300, expressed most strongly over parietal areas and linked to the detection of deviant or relevant stimuli. With the first signs of sleep oscillations—the theta rhythm—appearing in the EEG, cortical signal processing is reduced, which is reflected in the discontinuation of cortical potentials. Some authors assume that the cortical P300 component is replaced by a P450 with longer latency, appearing as a sleep-typical component following rare stimuli in NREM sleep. Similarly, cortical components such as the mismatch negativity (MMN), related to comparing new sensory stimuli to information in the sensory buffer, are also absent in NREM sleep. Deep, slow-wave sleep is furthermore associated with a lack of a contingent negative variation (CNV), which expresses learned expectancy of a second stimulus following a previous cue. Reactivity in sleep depends on stimulus characteristics—rare or strongly deviant stimuli induce K-complexes—whereas frequent or repetitive stimuli lead to a strong decrease or abolishment of most evoked potentials. To conclude, most components typical of information processing in wakefulness are absent in NREM sleep, but the altered mechanisms of processing during sleep still allow for two processes: the detection of strongly deviant stimuli or stimuli of personal relevance, as shown in responses to one’s own name during sleep.

During REM sleep, cortical evoked potentials bear more resemblance to wakefulness than to deep NREM sleep. Most late cortical potentials typical for wakefulness reappear in REM sleep, such as a P300 component, an expectancy-related CNV, or the MMN component. This suggests complex processing mechanisms, including reactivation of information stored in long-term memory. Usually these potentials appear with longer latencies than they have in wakefulness, reflecting delayed processing. In addition, the P300 component shifts toward a more parietal origin than during wakefulness, lacking contributions from frontal areas. However, stimuli do not consistently evoke responses in REM sleep, but intermittently during bursts of phasic activity, hardly induce any cortical reaction. Adolfo Llinas and Denis Paré have put forward the notion of the brain acting as a closed loop during REM sleep.

In general, regional metabolic or blood flow activity decreases in cortical areas during NREM sleep, starting at thalamic levels, and with a further prominent focus in frontal areas reflecting the predominance of slow EEG oscillations over frontocentral regions. In contrast, REM sleep is associated with regional increased metabolic or blood flow activity in deeper brain structures such as brain stem or thalamus, and in cingulate, limbic, and some cortical areas. Prefrontal and parietal areas display decreased activity throughout all sleep stages. Decreased reactivity to external stimulation is reflected in decreased activity of the corresponding processing centers during sleep, or even by a transient decrease in wider brain areas that may serve sleep-protective functions.

**Dream Mentation**

Whereas previous theories of dreaming such as those outlined by Sigmund Freud or Carl Gustav Jung have focused on the interpretation and intention of reported dream contents, a more recent focus
lies on formal, neuropsychological aspects and the neurobiological mechanisms underlying the physiology of dreaming, questions that can actually be targeted with hypothesis-based, scientific approaches. The proposal of Freud and psychoanalysts that dreams represent the fulfillment of wishes is challenged by the neurobiological-based conclusion of Allan Hobson, which argues that dream mentation results from brain stem–driven activation during REM sleep. Francis Crick proposed the hypothesis that we “dream to forget” to avoid information overload. Following the notion that dream mentation is not completely random but is modulated by waking experiences, hypotheses of dreaming as stabilization of emotional memory, simulation of threatening events, or individualization by genetic programming have been proposed, but clear empirical support is mainly lacking. Formal and content aspects of dreams are still a focus of research, which frequently tries to establish a link to personality traits of the dreamer.

Dream contents are usually forgotten unless sleepers wake up soon afterward. It can be assumed that dream contents remembered on awakening represent only a limited fraction of the REM mentations in the overall sleep period.

**NREM Sleep Mentation**

Sleepers, if asked whether “anything was going on in their mind” rather than “if they had been dreaming” will report some mental activity in about 50% of awakenings from NREM sleep. However, these mentations appear rather vague, representing rather a static scene or feeling, and usually lack the vividness found in “classic” dreams. Most research therefore has focused on REM sleep mentation.

**REM Sleep Mentation**

Sleepers report vivid dreams around 60% to 90% of the time if woken out of REM sleep. These dream mentations usually include actors, incidents, and a storyline. They often contain color and auditory imagery with hyperreal or hallucinatory qualities, whereas touch and taste sensations are rare, pain or smell sensations hardly ever occur, and gravity can sometimes be abolished. Voluntary control and self-awareness in dreams are reduced.

Hobson declared this state to be an enigmatic “third state” of consciousness in which, following his activation-synthesis model, the physiological brain stem–derived activation is synthesized to the—bizarre—dream mentation by sporadically activated higher brain centers in REM sleep. This important neurobiological theory of dream mentation and the ensuing focus on formal properties of dreams contests the so far prevailing interpretation of dream contents as found in psychoanalytical approaches. The vivid imagery of REM sleep-associated dreams is supposedly associated with increased activity in (secondary) visual cortex. Similarly, the intense emotions during dreams probably derive from activity in emotional centers of the brain, whereas the decreased dorsolateral prefrontal activity is held responsible for the bizarreness and incongruities of dream content. The precise role of frontal/prefrontal mechanisms and of neuropharmacology underlying dream mentations is still a matter of investigation. As activation during dreams is not independent of activation during wakefulness, damage to specific brain regions affects the respective characteristics of dreams.

**Nightmares**

The experience of nightmares, frightening dreams from which the person usually wakes up with recollection of the alarming dream content, is associated with increased heart rate and breathing frequency. Nightmares derive from REM sleep, whereas the similar night terrors occur without dream memories out of NREM sleep. Both are more frequent in early life. The high activity of the limbic system, especially the amygdala, during REM sleep is believed to induce the intense emotions associated with nightmares.

**Lucid Dreams**

Becoming aware of dreaming while dreaming is called *lucid dreaming*. During ordinary dreams, the sequence of events just happens, whereas in lucid dreams the dreamer can gain control of the course of dreams and can decide on actions. Persons experienced with lucid dreams can signal lucidity by volitional eye movements, enabling investigations of this state.

Renate Wehrle

See also Attention and Consciousness; Electrophysiological Studies of Mind; Neural Correlates of Consciousness
Further Readings


SMELL, PHILOSOPHICAL PERSPECTIVES

As far as the philosophical tradition would have us believe, vision is the model from which all theorizing about perception can proceed. Recently, however, philosophers have begun to reexamine tradition by turning their attention to the other modalities and what they, in their own right, have to tell us about the nature of perception. Still, discussions of olfactory experience are few. This entry introduces three broad areas of inquiry and discusses olfactory experience with respect to each. They are (a) the content of experience, (b) the nature of perceptual objects, and (c) the nature of the perceptual properties. Considering olfactory experience with respect to each presents unique challenges to the traditional visual-based model of theorizing about perception and, as a result, insight into perception unavailable from the visual case alone.

The Content of Olfactory Experience

It is commonplace to suppose that visual experience is world directed and, in particular, that it has representational content. We can think of the content of a perceptual experience as the way the world appears to a subject when she has that experience. If the world is that way, then the experience is accurate, or veridical. If it isn’t, then it is inaccurate, or nonveridical.

What must the world be like for an olfactory experience to be veridical? In other words, how does the world appear to be in an olfactory experience? Visual experience is importantly object based, presenting us with ordinary objects like apples and oranges. Perhaps what olfactory experience represents, then, is that there are ordinary objects in our environment with certain olfactory properties. There is no doubt that this view accords with the way we speak. Just as we say that the apple looks red, we say that the lilac smells sweet. According to William Lycan, olfactory experience does represent ordinary objects such as skunks and lilac blooms but only at a secondary level of representation. At a first level of representation, olfactory experience represents odors, or collections of airborne molecules. Lycan’s view also accords with everyday thought about the objects of olfactory experience. Not only do we say that the lilac smells sweet, we say that its smell lingers in the garden. And what we suggest when we say the latter is that olfactory experience also presents us with something in the air—a cloud or emanation of sorts.

Despite this, there remains reason to think that we speak more loosely about our olfactory experiences than we do our visual ones, in terms of their presenting ordinary objects as well as individual odors. In each case, the reason for suspicion draws on consideration of the phenomenology of olfactory experience. A characterization of the representational content of olfactory experience should honor the phenomenology of olfactory experience—that is, what olfactory experience is like. (Note that the phenomenological notion of content is not the only notion of content available, though it may be the most common notion at work in the philosophical literature; we will turn briefly to this issue below.) Taking the ordinary objects proposal first, consider a novel smell, one that you have no reason to suppose is the smell of one object as opposed to another. It is only once you know what the source of the smell is that you are able to make remarks such as “I smell the coffee.” This fact puts pressure on the view that olfactory experience represents ordinary objects. It
would seem that nothing in the olfactory experience itself “says” coffee.

There are further considerations against such a view—ones that avoid the controversial suggestion that kind properties (e.g., coffee) are eligible for representation in experience and that set aside issues of a perceiver’s capacity for identification. Many philosophers have held that object perception and, based on that, the ability for thought about individual objects, requires a robust form of spatial differentiation of the properties presented—for example, a figure-ground distinction. Vision achieves this; olfactory experience does not. Consider the experience you have when you indulge yourself in the smell of the breakfast cooking. (Note that smell denotes an olfactory property.) You are able to distinguish the coffee smell and the bacon smell, but your experience does not allow you to discriminate the particular objects that bear these properties—whether these are ordinary objects or odors (or both). Unlike visual experience, smells are not packaged together in space in such a way that these packages can be distinguished from one other and from a common ground. At any instant, it seems as if the smells are simply “here.” This not only puts pressure on the view that olfactory experience represents ordinary objects but also on the view that it represents individual odors.

Compared to the wealth of detail afforded by visual experience, then, olfactory experience seems a mere smudge. This “smudgy” feature of olfactory experience has prompted some philosophers to suggest that olfactory experience has no representational content. Subjectivist views of perceptual experience maintain that experiences are not world directed—that is, they have no “objective purport.” Subjectivist views are also characterized as the view that perceptual experiences are raw feels or “mere sensations.” Although discussions of olfaction in the philosophical literature are rare, the subjectivist view is held up as the prima facie view of the nature of olfactory experience.

But a subjectivist view is not inevitable. There is a moderate representational view available that honors the phenomenology of olfactory experience and yet maintains that it is world directed. Recently, discussions of the representational content of visual experience have focused on a debate between the view that visual experience has existentially quantified, or abstract, content and the view that it has object-involving content. To see the difference between these two views, consider the visual experience you have when you look at an orange on the counter. According to the existentially quantified account (from here on, the abstract account), your experience has the following content: There is some object $x$ at location L and $x$ is orange, oval, and so on. According to abstract theorists, it is possible that experiences of two qualitatively identical, yet distinct, oranges might be phenomenally indistinguishable. Indeed, a perceiver might hallucinate an orange before her and yet have that experience be phenomenally indistinguishable from a veridical experience of an orange. To preserve this possibility, the abstract theorist proposes that the content of each is content into which no particular object enters. The object-involving theorist, on the other hand, claims that such a view ignores the particularity of visual experience. It’s not that some object appears to be orange, oval, and so on. This one does! The very orange before you, then, must be a part of the content of your experience. Letting $o$ name the actual orange before you, the object-involving account claims that your experience has the content: $o$ is orange, oval, and so on . . . and at L.

This entry will not consider the solution to this debate about visual experience. The debate itself, however, draws attention to the moderate view of olfactory content. Given its phenomenology, an object-involving account is unsuitable for olfactory experience. Unlike visual experience, olfactory experience does not seem to present particular things. As a result, olfactory experience cannot live up to the particularity that such a view demands of experience. However, the abstract account, which requires no such particularity, seems like a natural fit. A version of it can respect that smells are experienced as external to a perceiver and that olfactory experience does not present us with the individual objects that instantiate those smells. Drawing on the considerations of spatial presentation discussed above, the moderate view posits that olfactory experience only ever represents that a single “something or other” is smelly and “here.” On this view, the following schema specifies the content of any olfactory experience: There is some $x$ here that is $F$, $G$, and so on. If there is nothing in the vicinity that is $F$ and $G$, then the experience is nonveridical.

There are several reasons to think that any world-directed view is preferable to a subjectivist one.
Despite their difference in phenomenology, we still think of the senses as informational systems. Using the senses, we are able to gather information about the world. Although we might think their phenomenology is impoverished, our olfactory experiences still function to guide behavior and action. If someone smells smoke in the building, they flee. As guides of behavior and grounds of belief, the experiences of the sense modalities form a common kind. A shared world directedness provides a way of accounting for this commonality.

Olfactory Objects

If we accept the abstract view, the next question is, given that some olfactory experiences are veridical, what objects have the properties those experiences present? That is, what are the olfactory objects? As we have seen already, there are several options: Olfactory objects are (1) “source” objects, (2) odors, or (3) both odors and source objects.

Given the traditional, “visuocentric,” approach to theorizing about perception, we might feel tempted to say that the properties of which we are aware in olfactory experience are qualities had by regular old objects—lilac blooms, skunks, and pots of coffee. Although olfactory experience is not discriminating enough to report that there are particular objects, on Proposal 1 the bearers of the properties presented in olfactory experience are in fact ordinary objects. We certainly think of lilacs, skunks, and portions of coffee as the sources of smells. But we also speak of them as having a good or bad smell. Consider how, when rooting around in the fridge for the rotten food, you say of the uncovered cabbage, “It’s this that stinks.” What you direct our attention to is the head of cabbage. Or to take a more pleasant example, we take pride in the roses in our garden, not only because they look beautiful but also because they have wonderful smells. We attribute a property—namely, a smell—to the rose. The same is true of the cabbage. It has a very bad smell, we say.

This might seem a common view, but it is also subject to question. Consider how you can have an olfactory experience even though the object that you think of as responsible for the smell is far away. For example, you might smell the rubbish from your apartment window even though it is outside in the bin. If my olfactory experience represents that properties are instantiated by something or other “here” then your experience must be nonveridical. The rubbish is not anywhere near you; it is downstairs and outside. The problem with Proposal 1 arises because this kind of circumstance is not rare. If olfactory objects are things like piles of rubbish, many of our olfactory experiences will turn out to be nonveridical. And this is a view that we ought to avoid.

Proposal 2 has it that olfactory objects are collections of airborne molecules given off by the rubbish—that is, odors. On this proposal, the experience you have when you smell the rubbish through the window is a veridical one. The rubbish gives off an odor, that odor is “at” you (indeed, it has gone up your nose), and your experience reports it as such. This is a more plausible result. Experiences we intuitively count as veridical turn out to be so.

Nevertheless, we might feel drawn to the view that the rubbish also has the stinky property. After all, the rubbish gives off or emits the odor. If we feel the pull of such a view, there are two ways that we can accommodate it. First, we might hold that the rubbish has the stinky property but that olfactory experience does not present that object (i.e., it presents the rubbish odor only). The second option is more controversial. We do think of ourselves as smelling the rubbish by smelling the odor it emits. If we take this proposal seriously, as denoting something about the content of olfactory experience, then we arrive at Proposal 3: Olfactory objects are both odors and source objects.

Lycan’s multilayered view of representational content, noted above, is a version of Proposal 3. Both odors and ordinary objects are olfactory objects, each definitive of successive levels of representation. According to Lycan, insofar as we might think that there is something both correct and incorrect about a situation in which a rose odor is present when no rose is, his view respects our judgments. On one level, one’s experience is veridical; on another, it is nonveridical.

But the view that olfactory objects are (also) things like roses faces a challenge. If ordinary objects are among the olfactory objects, then content cannot be determined by phenomenology alone. As we have seen, the phenomenological notion of content and, in particular, the abstract view applies nicely to Lycan’s first level of representation (where the represented objects are odors); it does not, however, fit the second (where the represented objects are ordinary objects). Advocates of such a view,
like Lycan, owe us an account of an additional kind of content and an argument for why we ought to think that olfactory experiences have that kind of content. One candidate view is the teleological view of content advocated by both Ruth Millikan and Fred Dretske—in short, the view that the content of an experience depends on its function within the system, or organism, of which it is a part. This, it would seem, is just the approach that Lycan takes—although he endorses it not only for the second level of representation but also for the first.

**Olfactory Properties**

A final issue to consider is the nature of the smells themselves. Although very little has been written about olfaction in the philosophical literature, the little that there is reveals a contrast in the views favored for the cases of smell and color. For many philosophers, the view that colors can be explained in purely physical terms has seemed very appealing. In the case of smell, this kind of view has seemed less appealing. Those who have discussed olfaction have favored either dispositionalism or projectivism. According to dispositionalism, smells are dispositions to cause certain kinds of experiences in perceivers. Dispositionalists do not deny that smells are properties of objects (we will assume, for the sake of discussion, that these objects are odors), but they do maintain that the nature of these properties cannot be specified without reference to experience. According to dispositionalists, the lilac smell, for example, is the disposition to cause a distinctive kind of experience in suitable perceivers. Unlike dispositionalists, projectivists argue that the lilac smell is a mental property and that something internal to the perceiver (e.g., a sense datum, experience, or portion of the “olfactory field”) has that property. Projectivists argue that these properties are then “projected” onto objects in the external world.

Although dispositionalism and projectivism each take into account the “felt character” of olfactory experience, they fall on two sides of one debate over the nature of what John Locke called the secondary qualities—colors, sounds, smells, tastes, and feels. This is the debate between realism and eliminativism. Olfactory realism claims that things in the world have the property ascribed to them by olfactory experience. Dispositionalism is just such a view. Eliminativism, on the other hand, claims that smells are not properties of things in the world. In claiming that smells are properties of experiences, projectivism is a form of eliminativism.

Because it renders all olfactory experience illusory, eliminativism is often regarded as a last resort, a kind of view rendered plausible by the failure of any realist view. Realism, then, is the default position for any view of the secondary qualities. All realist positions fall into one of two camps: relationalism and nonrelationalism. According to relationalism, smells are constituted by relations between objects and perceivers. Dispositionalism is one such view. Nonrelationalism, on the other hand, maintains that smells are properties independent of the perceiver or mind. A natural nonrelationalist position is that smells are molecular properties (also known as physicalism).

There can be significant intersubjective differences between the ways that perceivers smell certain odorants to be and, on these grounds, relationalists can argue against nonrelationalism as follows: Significant intersubjective differences in perceived smell exist. If smells are nonrelational properties of odors, then the smell of an odor does not in any way depend on the experiences that perceivers have when they come into contact with it. So there ought to be a unique smell that an odor has and, as a result, a reason to favor one group of perceivers over any other—that is, those whose experiences present the smell of the odor. But there is no reason to favor one set of perceivers over another. Because it cannot live up to its own demands, nonrelationalism is false.

As it turns out, the issue is not as simple as this argument might make it initially appear. Nonrelationalism is threatened only if these intersubjective differences involve major shifts in perceived quality. If the differences involve minor shifts in perceived quality then the threat is little to none. All sensory systems have limits of resolution. And when a system is pushed to the limits of its resolution, it is bound to make minor mistakes. If shifts in perceived quality are the result of just such a mistake then we do have a reason to favor one group of perceivers over another. One group is simply mistaken—that is, their experiences are nonveridical.

One opportunity for future research on the nature of smells, then, involves determining whether the shifts in perceived quality are major or minor. But we are yet to have a structured quality space for smell and, as a result, lack a model by which
to evaluate differences in perceived quality. Many systems have been proposed; none has been found satisfactory. Given the breadth of our olfactory discrimination, each system has been accused of oversimplifying olfactory experience. Obviously some sort of consensus is necessary before we can evaluate claims of perceptual variation. What we can be sure of at this point is that future philosophical work on the olfactory properties will progress with further developments in olfactory psychophysics.

Clare Batty

See also Disjunctive Theory of Perception; Hearing, Philosophical Perspectives; Intentionality of Bodily Sensation; Realism and Instrumentalism; Taste, Philosophical Perspectives

Further Readings


Social Cognition

Although researchers in many different sciences use social cognition to refer to a wide range of phenomena, the present entry uses the term to refer to an enormous body of theoretical and empirical work that arguably is the largest and most dominant area within social psychology. But this was not always so, and it is instructive to trace social cognition research back to its beginnings in the late 1970s. At that time, social cognition was a label referring to research conducted by a small group of social psychologists who were interested in how people form impressions of others, or “person memory.” Although social psychologists had long been interested in how people form impressions of others, they tended not to focus explicitly on cognitive processes such as how information is encoded, stored, and retrieved from memory. The original social cognition researchers had this explicit focus, thereby distinguishing themselves from other social psychologists, including those who also were interested in impressions of others. The explicit focus on cognitive processes resulted in the development of detailed person memory models that made specific predictions. Additional successes followed, and social cognition became an important area of social psychology.

Person Memory: The Original Social Cognition

Imagine an observer who notices the behaviors of the person who is the target of her observations (i.e., the target person). In addition, suppose the observer has a prior expectancy about the target person and subsequently observes the target person’s behaviors that are congruent or incongruent with that prior expectancy. How will the target person’s behaviors become represented in the observer’s memory? The dominant cognitive view was embodied in the notion of a “schema,” which is a cognitive structure that tends to preserve information consistent with the schema and bias recall and judgments in that direction. If prior expectancies act as schemas, then the schema should preserve information that is congruent with it in memory and later recall of the target person’s behaviors should be biased in the direction of the schema. For example, if the observer had a prior expectancy that the target person is kind, then the target person’s kind behaviors should be better recalled than her unkind ones. In contrast, early social cognition researchers suggested that perhaps incongruent information is more informative than congruent information, and so people form more associations concerning incongruent behaviors than congruent ones. The data supported this reasoning.
As more information accumulated in the early 1980s, person memory theories became more detailed. An important book by Robert Wyer Jr. and Thomas K. Srull titled *Memory and Cognition in Its Social Context* summarizes the theories and empirical findings. In general, the “received view” is that there are two levels at which information about people is represented in the observer’s memory, and the representations at both levels are in the form of “nodes” or concepts. The higher level has a “higher order person node,” which contains the observer’s prior expectation and other general information about the target person. The lower level has congruent and incongruent behavior nodes (as we have already seen) and irrelevant behavior nodes that have nothing to do with the prior expectancy. For example, the target person’s eating a sandwich is irrelevant to the observer’s prior expectancy concerning the target person’s kindness. The three lower nodes representing congruent, incongruent, and irrelevant behaviors are associated with the higher order person node by virtue of the fact that the target person has previously performed these behaviors. The strongest of these three associations is between the higher order person node and the congruent behavior nodes.

But there are other ways in which associations among the behavior nodes are formed (see Figure 1). Because congruent or irrelevant behaviors are easy to understand and do not cause the observer to question her expectancy, they do not stimulate the formation of associations between behavior nodes. In contrast, the performance of an incongruent behavior requires far more processing. To understand the target person’s incongruous behavior, the observer compares it to other incongruent or congruent behaviors. Each time the observer makes such a comparison between behaviors, she forms associations between the nodes representing them. Thus, she forms associations between nodes representing incongruent behaviors and other incongruent behaviors, and between incongruent behaviors and congruent ones, but not between nodes representing congruent behaviors and other congruent ones, all while isolating irrelevant item nodes in the network. As a result, there are more associative pathways leading to the recall of incongruent behaviors (from incongruent and congruent ones) than to the recall of congruent behaviors (only from incongruent ones) and fewest leading to the recall of irrelevant behaviors.

A number of predictions can be derived from the received view. Most obviously, recall should be greatest for incongruent items and least for irrelevant ones. Less obviously, when an incongruent item is recalled, there are associations leading to both congruent and incongruent items, and so the next item recalled is likely to be a congruent or an incongruent one. But when a congruent item is recalled, then all the inter-item associations lead to incongruent items, and so the next item recalled is likely to be an incongruent one. When irrelevant items are recalled, they are associated only with the higher order person node, which, in turn, is more highly associated with congruent than incongruent items. Thus, the theory predicts that when people recall an irrelevant item, they should traverse an associative pathway first to the higher order person node and then to a congruent item node, thereby leading to the recall of the congruent item. Experiments confirmed all these predictions as well as some others not mentioned here. It is important to appreciate the emphasis that went into encoding (specifying the process by which associations are formed), storage (the structure actually stored in memory), and retrieval (how people traverse associative pathways based on how they were stored), and how this careful theorizing led to interesting predictions.

Although further work in the person memory area continued, the frequency of these investigations decreased in the 1990s, and person memory research is rare today. Although many researchers currently
Social Cognition

706

consider person memory to be a “dead” area, its historical importance should be appreciated. Person memory gave social cognition its initial impetus, and there is no way to know whether social cognition would exist at all, or in what form, had the early researchers chosen to study something else.

Category Accessibility: The Other Early Social Cognition Area

Imagine you had been told about a person named Donald who crossed the Atlantic Ocean in a sailboat. Would you judge Donald to be “adventurous” or “reckless”? Early social cognition theorists suggested that much would depend on what other information was accessible to you. If the category of “adventurous” happened to be accessible then you might be likely to encode the behavior as an exemplar of that category, thereby implying that it was adventurous, but if the category of “reckless” happened to be accessible then you would probably encode the behavior as an exemplar of recklessness, leading you to the judgment that it was reckless. In fact, early research demonstrated that manipulating the categories accessible to participants in experiments (i.e., “priming” these categories) caused them to make judgments in the direction of whatever category was primed.

The notion of priming categories to influence judgments expanded in the 1980s in two important directions. Specifically, it became clear that categories could be primed more or less recently (just prior to judgments or well before judgments) or more or less frequently (many times or fewer times prior to judgment). The impact of recently (and nonfrequently) primed categories tends to be strong immediately and decrease rapidly over time, whereas the impact of frequently primed categories tends to decrease more slowly over time. Two sorts of theories account for the effects. According to the bin model, information is accessed depending both on the number of times it is represented in the bin and where it is represented in the bin. Thus, in the case of a single recent prime, this information is sitting at the top of the bin, there nevertheless remains a reasonable probability that one of the category representations will be used for making judgments.

An alternative account invoked the concept of decay of excitation. The more frequently a concept has been primed, the slower the rate of decay of its excitation, which is another way of saying that the concept retains much of its ability to influence judgments over time. As a matter of history, although neither of these general theories is particularly influential today, the notion of category accessibility and the empirical technique of employing priming manipulations have been incorporated by almost every subarea within social cognition. It is probably not an exaggeration to say that category accessibility and priming are so ingrained into the social cognition lore that they currently are considered as givens.

Social Cognition, Amalgamation, and Contemporary Themes

Throughout the 1980s, the person memory and category accessibility domains gained in popularity and academic acclaim. Researchers in other domains began to associate themselves with the increasingly trendy area of social cognition, and areas of social psychology that were not considered originally to be within the social cognition domain became so categorized. For example, although attribution theories had been cognitive from their beginnings, it was not until the mid-1980s that they were considered to be part of social cognition. Likewise, the area of attitudes became associated with social cognition. This expansion of social cognition applied eventually to almost every area of social psychology, to the point where it would be difficult to find a social psychology area that has not been influenced by social cognition theories, methodologies, or both. In addition, this expansion has resulted in theories that assume two processes (dual-process theories) that can be distinguished from theories that do not assume two processes.

Dual-Process Approaches

In contemporary research, there has been a great deal of emphasis on the distinction between mental processes that are automatic versus controlled, to the point where the vast majority of 21st-century social cognition theories are dual-process theories.
Put simply, these theories posit an automatic process that does not use up cognitive resources and does not involve conscious volition and a controlled process that uses cognitive resources and does involve conscious volition. Examples of dual-process theories include that persuasion happens via a peripheral (automatic) or a central (controlled) route, or that behaviors result from habitual (automatic) or non-habitual (controlled) processes, or that the formation of plans for implementing behavioral intentions makes automatic the intentional process that otherwise would be controlled, and so on.

New methodological paradigms have been developed to exploit the distinction between automatic and controlled processes featured in these theories. Researchers often collect reaction time data, under the assumption that if two cognitions are associated, priming one of them should increase the speed with which the other will be accessed. Thus, in a study on prejudice, participants might be primed on some trials with a word toward which they have a positive attitude and some words toward which they have a negative attitude. Either way, after being primed, the participants are asked to respond to the names of black or white persons. Those participants who are prejudiced against black people should respond more quickly to a positive prime followed by the name of a white person or to a negative prime followed by the name of a black person, than to positive-black or negative-white combinations.

Another widely used social cognition paradigm is subliminal priming. The idea is as follows. If a hypothesized social cognition process is automatic and conscious volition is not required then relevant behaviors can be influenced outside of consciousness with subliminal priming. For example, priming people at a subliminal level with pictures of elderly people caused them to walk more slowly than those who had not been so primed.

Notwithstanding the popularity of dual-process approaches, they are not touted by everyone. Although different cognitive processes may use up different pools of resources and be automatic with respect to each other, some would argue that few (if any) of the cognitive processes of interest to social psychologists use up exactly zero resources. If so, there would be very few (if any) absolutely automatic processes; on the contrary, one would expect a continuum of resource utilization, and so the notion of separating social psychology phenomena into dichotomous automatic versus controlled groupings would not make sense.

**Other Social Cognition Approaches**

A substantial amount of social cognition work does not use dual-process theories. An example is the work of cross-cultural psychologists who have distinguished between people in individualist cultures who emphasize private self-cognitions (i.e., their traits and states), as opposed to people in collectivist cultures who emphasize collective self-cognitions (i.e., their group memberships). Social cognition thinking suggested two opposing ways in which self-cognitions might be organized. One possibility is that there is only one self-concept but that culture influences the relative proportions of different types of cognitions associated with it; individualists have more private self-cognitions than do collectivists and collectivists have more collective self-cognitions than do individualists. A second possibility is that there are different cognitive structures for storing different kinds of self-cognitions; private self-cognitions are stored in the private self and collective self-cognitions are stored in the collective self. If so, culture would influence the relative accessibility of these two cognitive structures.

The different possibilities make opposing predictions that can be tested empirically. According to the two-selves theory, priming the private self should increase the retrieval of private self-cognitions, whereas priming the collective self should increase the retrieval of collective self-cognitions. But if there is only one self-concept with no internal organization (if there is internal organization, then this turns into a multiple self-theory), priming to differentially influence the retrieval of private versus collective self-cognitions should make no difference. A second prediction concerns the order in which self-cognitions are retrieved. The two-selves theory indicates that private self-cognitions are associated with each other but not with collective ones, and collective self-cognitions are associated with each other but not with private ones. If so, then the retrieval of private self-cognitions should be likely to be preceded by the recall of private ones, and the recall of collective self-cognitions should be likely to be preceded by the recall of collective ones. But if the one self-theory is true, retrieval should be random. Findings concerning both predictions came out in favor of
the theory that there are at least two self-concepts. More recent research has expanded this theorizing to assert that priming different self-concepts influences the process of thinking as well as the content of the thoughts.

Social cognition also has been used to test for the internal structure of attitudes—whether attitudes have both cognitive and affective components. In one experiment, participants wrote down their attitudinal beliefs, and blind coders rated them as mostly cognitive or mostly affective. The researchers hypothesized that during the process of forming an attitude, cognitive beliefs are compared to each other to form the cognitive component, and affective beliefs are compared to each other to form the affective component, but there is little reason for participants to compare cognitive and affective beliefs. Consequently, the attitude formation process should cause stored associations between cognitive beliefs and other cognitive beliefs, between affective beliefs and other affective beliefs, but not between cognitive and affective beliefs. During retrieval, participants should have been able to traverse associative pathways between cognitive beliefs and other cognitive beliefs, between affective beliefs and other affective beliefs, but not between cognitive and affective beliefs. This led to a prediction that retrieval would be clustered by belief type (cognitive beliefs retrieved adjacently to each other and affective beliefs retrieved adjacent to each other). Experimental confirmation of this prediction supported the hypothesis that attitudes have separate cognitive and affective components.

Conclusion

Social cognition has been a major force in social psychology and continues to expand, both theoretically and methodologically. It is very different today than it was at its conception, and it may metamorphose into new forms that would be unrecognizable to its originators. There is no way to predict which new areas of psychology social cognition will influence, but social cognition researchers can look forward to exciting research yet to come.

David Trafimow

See also Attitudes and Behavior; Attribution Theory; Persuasion; Two System Models of Reasoning

Further Readings


Social loafing is a broad construct that includes the free rider effect and the sucker effect. The free rider effect refers to the decline in contributions observed when people perceive their contributions as redundant with the contributions of others. The sucker effect is the tendency for people to withhold contributions when they perceive or anticipate that other people are loafing at their expense. That is, people sometimes contribute less to avoid being exploited by others who are loafing. Social loafing is distinct from these other effects in that social loafing is a broader term referring to any reduction in effort or motivation when contributions are pooled. The free rider effect and the sucker effect are narrower terms that refer to specific causes of social loafing.

Researchers have documented social loafing in a variety of tasks, including clapping and cheering, swimming in a relay race, solving mazes, wrapping gum, pulling a tug-of-war rope, and generating uses for objects. In a typical study, groups of participants receive instructions to generate uses for an object (e.g., uses for a knife) and believe that their individual contributions are or are not identifiable. Participants reliably generate more uses when they believe the number of uses they generate can be identified by the experimenter than when they believe the uses cannot be identified.

**Causes of Social Loafing**

Group performance tasks can be regarded as having three components: (a) an effort expectancy component (i.e., expectations regarding the contributions group members make), (b) a performance expectancy component (i.e., expectations regarding the product of those contributions), and (c) an outcome value component (the consequences such as the reward or punishment that result from the performance). Research links social loafing to each component. Regarding effort expectancy, people are more likely to loaf if they perceive that their individual contributions are irrelevant to achieve a good group performance. For example, if a man assembling widgets in a factory believes that he and his coworkers will meet the day’s production quota regardless of whether he personally works hard or loaf, he is likely to loaf. Similarly, if the man believes that the group will not meet the day’s quota regardless of how hard he works, then he also is likely to loaf. In both cases, reaching (or not reaching) the quota is irrelevant to the man’s individual efforts. Research documenting the role of effort expectancy in social loafing comes from studies that manipulate the redundancy of individual efforts. These studies show that people are more likely to loaf when they perceive their efforts to be redundant with the contributions of other group members than when they perceive their efforts as unique.

Regarding performance expectancy, people are more likely to loaf if they perceive that the quality or quantity of the group’s performance is unrelated to the outcome. For example, if the factory worker believes that his work group will receive the same bonus regardless of whether the group meets the production quota, he will loaf. Similarly, if the worker believes that his work group will receive no bonus (and no punishment) regardless of whether the group meets the quota, he will loaf. In short, people in a group will work hard when they believe rewards and punishments are linked to the group’s performance and will loaf when they perceive no such link. Evidence linking social loafing to the performance expectancy comes from studies showing that people loaf when they perceive that there is a low likelihood that a good group performance will be rewarded. In addition, the finding that people loaf when group member contributions cannot be identified supports the role of the performance stage in social loafing. When contributions cannot be identified, group members receive no personal reward for working hard or punishment for loafing.

Regarding outcome, people are likely to loaf to the extent that the value of a good group performance is low. Value is low when there is no reward for achieving a good group performance, when the reward is not valued, or when there are overriding costs. For example, the factory worker may perceive that the bonus for making the day’s production quota is too small to justify working hard. That is, the cost of working hard exceeds the benefits. Research on outcome value demonstrates that offering powerful incentives for a good group performance reduces social loafing. For example, one study showed that participants wrapping gum in groups wrapped more if they believe the gum they wrapped would be donated to a worthy cause (an intrinsic benefit) than if they believed it would not.
Reducing Social Loafing

According to theorists, effort expectancy, performance expectancy, and outcome value must all be high or people will loaf. That is, if any one of these factors is low then people will reduce their contributions to the group. Yet social loafing is not inevitable. People who depend on groups to accomplish tasks as well as the group members themselves can reduce social loafing by making sure three conditions are met. First, group members must perceive that a good group performance depends on their individual contribution. That is, group members must perceive that if they personally withhold contributions, the group performance will suffer. Second, group members must perceive that a good group performance will be rewarded and that a poor group performance will not. The surest way to create this condition is to make individual contributions identifiable. Third, there must be an adequate reward or incentive for achieving a good group performance, one that exceeds the cost of contributing.

James A. Shepperd and Darya Melnyk

See also Collective Action; Group Decision Making

Further Readings


Spacing Effect

Practice that is distributed over time rather than massed together within a shorter period has been shown to have a stronger effect on later performance. This spacing effect is an improvement in future memory retrieval performance caused by more widely spaced practice (where repeated learning events are distributed temporally rather than massed at a single time). Research has shown that this effect is very robust for the learning of factual information such as words, images, definitions, or artificial stimuli such as random letter strings and that this benefit persists over delays as long as several years. The spacing effect was first recognized by Hermann Ebbinghaus in his study of the learning of serially ordered lists of words. He found that the massing of study on a single day resulted in poor performance when attempting to relearn items as compared to faster relearning when the study of the word lists was distributed over a 3-day period. This entry briefly describes the spacing effect and discusses three types of theories that attempt to explain it.

Features of the Spacing Effect

The spacing effect is well established in memory-dependent learning tasks and has been demonstrated most frequently using paired-associate and free-recall memory tasks. In a paired-associate spacing-effect experiment, pairs of words or other stimuli are associated over repeated practices that are spaced in time by practicing other item pairs during spacing intervals. Recall in paired-associate memory tasks may be tested at the conclusion of practice by prompting with one member of each pair to cue recall of the other pair member, but many experiments include a buffer task or a short-term or long-term interval (typically at least a day) between learning and retention, since a paired-associate spacing effect seems to require a significant retention interval to become apparent.

In a free-recall task, participants learn a word list in which some of the words repeat, again with spacing defined by the number of intervening items between repetitions. The effect of spaced practice of items in a list is determined with testing after the conclusion of practice, either immediately or with some retention delay interval. Spacing effects in free-recall lists are often apparent without the need for a significant retention interval.

While the spacing effect has been most thoroughly studied in memory tasks and might be expected to occur in any task with a memory component, it has also been reported in motor skill tasks. However, there has been much less research on the spacing effect in motor skill tasks, and it is still unclear to what extent spacing effects are universal to all types of learning. In fact, many researchers have proposed multiple mechanisms for spacing effects, and it is possible that spacing effects occur because of different mechanisms depending on the task.

The spacing effect often shows strong interactions with the amount of practice and the duration of the retention interval. The practice by spacing
interaction is characterized by an increasing benefit from spacing as practices accumulate. This interaction means that additional spaced practices beyond the second trial will continue to contribute more learning if the practices are widely spaced. The spacing by retention interval interaction, which can only be detected using experiments with multiple retention intervals, is a tendency for much more rapid forgetting following practice that is narrowly spaced (i.e., the effect of spacing is greater after longer retention intervals).

**Theories of the Spacing Effect**

The history of spacing-effect research is rich with suggested explanations for the effect.

**Fluctuation Theories**

Fluctuation theories propose that each repetition of an item to be memorized results in a sample of the item and its context features (stimulus components) being committed to memory. Learning occurs through the accumulation of these samples, and retrieval is determined by how well the components of a later cue match the stimulus components learned. In a fluctuation theory, the stimulus components are supposed to fluctuate with the passage of time. Because of this fluctuation, more widely spaced practice may create an overall encoding that includes a broader share of the possible stimulus components, while narrow spacing results in a rather redundant encoding because very little fluctuation of the available components can happen between repetitions. Two main varieties of fluctuation theory have been proposed: contextual variability, which focuses on the importance of the entire context in building a broad representation, and encoding variability, which focuses on the central importance of fluctuations in the sampling of the items rather than the broader context.

**Accessibility Theories**

Typically considered as an alternative to fluctuation theories, another class of theory focuses on how the current memory accessibility at the time of repetition results in different effects for each repetition. These “accessibility” theories describe how narrow spacing results in memories that are more accessible at the time of repetition, but they propose that it is this accessibility at repetition that reduces further learning when the repetitions are narrowly spaced. Earlier versions of accessibility theories described how priming or habituation from narrow spacing might block learning; these explanations are similar to the idea of “desirable difficulty” popularized by Robert Bjork because wider spacing increases difficulty by reducing accessibility. Later versions of accessibility theories have elaborated this theory by describing how increased accessibility at repetition may lead to learning that is forgotten more quickly as opposed to learned more slowly. Like fluctuation theories, the newer accessibility theories explain the spacing by retention-interval interaction (the increased benefit from spacing after long retention intervals), which older accessibility theories did not attempt to explain. These newer accessibility theories map well to neurophysiology research, which shows that neurons exhibit this same tendency to show reduced long-term effects from narrowly spaced stimulation. Similar to accessibility theory, the study phase retrieval theory of spacing effects says that when a narrow repetition occurs before the previous repetition has left working memory, learning will be very poor.

**Strategic Theories**

Though paired-associate spacing effects are usually explained with one of the theories above, strategic processes such as differential rehearsal (rehearsing spaced items more) and rehearsal organization (rehearsing clusters of items) have also been shown to produce spacing effects in free-recall list-learning experiments. Strategic explanations are typically proposed only for list-learning spacing results.

*Philip I. Pavlik Jr.*

**See also** Desirable Difficulties Perspective on Learning; Rehearsal and Memory; Skill Learning, Enhancement of; Spacing Effect, Practical Applications; Working Memory

**Further Readings**


Spacing Effect, Practical Applications

Placing a temporal gap between study sessions increases the amount of information remembered in the future, a phenomenon called the spacing effect. The spacing effect allows students and teachers to efficiently allocate a fixed amount of study time to maximize later retention of a set of facts or a new skill. Educational environments that may benefit from spacing include completion of homework assignments, learning within a classroom setting, studying for exams, business training courses, computer-aided learning, and language-learning courses. Real-world tutors that implement spaced study include SuperMemo, the Mnemosyne project, and the Pimsleur language learning system.

Since the spacing effect was first described in 1885 by Hermann Ebbinghaus, hundreds of studies have examined its effects using materials with classroom and real-life utility. Related to school curriculum, this includes vocabulary, fact, and prose memorization; related to work and leisure activities, it includes typing, tossing balls, and playing video games. From these studies, a set of concrete recommendations for applying the spacing effect to real-world learning can be made.

Practical Techniques

In general, a long temporal gap between learning sessions increases the level of future recall. While very long gaps of at least a month are necessary to produce long-lasting benefits to retention for both verbal materials and motor tasks, relearning on a daily basis is more effective if the learner cares solely about performance on an upcoming exam or performance. Taking a few minutes break between each study session is always preferable to learning material within a single, massed study period. In general, too little spacing between study sessions is quite harmful to retention while too much spacing only leads to small decreases in later recall. An optimal level of spacing can more than double later recall.

The use of cumulative exams is perhaps the simplest technique teachers can implement in their classroom if they wish to promote durable memory for the material being taught. Cumulative exams encourage students to learn the same set of material on at least two separate occasions, providing the essential temporal gap between each learning episode. Other useful instructional design choices include systematic quizzes on previous topics, use of classroom time to review key curriculum content, and homework assignments that emphasize the primary points to be learned. Homework assignments and quizzes that mix, shuffle, or interleave different topics, such as addition and subtraction, have been shown to improve later performance.

Because testing with feedback is more beneficial than restudy alone, students should learn using flashcards rather than merely by (re)reading textbook chapters. When testing with feedback is used during study sessions rather than restudy alone, benefit from study episode spacing is substantially increased. After an exam, teachers should provide delayed feedback about the correct answers, which provides a form of spaced learning. Ideally, this should involve sequentially presenting a frequently missed test item, asking students to generate an answer to that item, and then providing feedback about the correct answer. Unless feedback is provided, students are quite unlikely to fix misconceptions about the correct answer. Tests and exams should emphasize key points rather than minor details so that students retain the most important material being taught. More key material will be remembered if fewer superfluous details are presented within textbook chapters and during classroom lectures. This instructional design choice allows educators to devote more time to spaced restudy of key points.

Areas Needing Further Study

While spacing effects are robust for most forms of verbal and skill learning, using both recall and recognition memory measures, such effects have not been found consistently for all types of learning. Category induction, such as learning to identify whether a novel skin lesion is benign or cancerous, showed spacing benefits in one study while another study failed to
find a spacing advantage. Virtually all research has used accuracy rather than speed as an outcome variable. Thus, effects of spacing on speed of task performance are unknown. Likewise, effects on higher level, critical-thinking skills, such as synthesis of material and creation of novel ideas, are unknown.

Some research has addressed the use of increasing gaps between a series of several study sessions, instead of a fixed gap, and this increasing-gap approach is used in most commercial-tutorial software. However, research that examines effects of fixed versus increasing gaps is inconclusive. While spacing benefits have been demonstrated from infancy through older adulthood and across a wide variety of ethnic groups, it remains unknown whether different age groups require different spacing intervals to optimize retention.

Nonintuitive Aspects

Because short gaps between study episodes lead to higher immediate recall, teachers and students may incorrectly feel they should dispense with recommendations to implement spacing in their classroom or in their study habits. Because the goal of education is to provide students with a body of knowledge and a set of skills that follows them throughout life, this intuition is misguided. It is easy for students to be misled into thinking they are learning efficiently and mastering a set of material, when in fact they are not. Educators and students should remain cognizant of the fact that worse immediate recall can mean better long-term recall.

Nicholas J. Cepeda

See also Desirable Difficulties Perspective on Learning; Rehearsal and Memory; Retrieval Practice (Testing) Effect; Skill Learning, Enhancement of; Spacing Effect

Further Readings


Spatial Cognition, Development of

Spatial ability is necessary to much of human activity. Human adults are highly skilled at many forms of spatial skills. We can all find our car in the vast parking lot and find our way home after the game, but where does this powerful and essential skill come from? How do we develop the ability to explore our world and still find our way home? This entry discusses the development of spatial cognition, first by discussing some of the theoretical issues that frame current debates in the research. Next, it presents an overview of the spatial system that emerges and, lastly, it provides a brief sketch of the changes in spatial ability over developmental time.

Theoretical Issues

Spatial ability has been an area of fierce debate in regard to the nature and course of cognitive development. Data-driven arguments are made from diverse perspectives as we gain information about the specifics of the changes in the spatial system over developmental time. Although much of the current research can trace its roots to the writing of Jean Piaget, the actual claims of a Piagetian view have come under increasing fire. There are a number of competing views vying to offer a new view of the development of spatial cognition. The first of these perspectives comes out of the repeated finding that Piaget underestimated the abilities of young infants. This view espouses a nativism (the claim that most if not all cognitive ability is inborn) that puts much emphasis on early ability and infant competence and less stress on later developments. A second view comes out of an interest in Vygotskian social learning (a view that most ability emerges from social experience and active tutoring) and puts emphasis on spatial language and the cultural milieu surrounding spatial development. The third view is an interactionist perspective that represents an attempt to integrate nativism and constructivism (the Piagetian view that children construct their own cognitive structures through active exploration) into a theory that accounts for early competence and subsequent developmental change. Interactionist models generally embrace the idea that cultural issues and environmental influences as well as early starting points
and maturation all combine to produce an individual's developmental trajectory.

**The Spatial Coding System**

The development of spatial cognition starts out with a set of primitive responses that rapidly become the complex system that makes a toddler able to help us find our keys in the morning. Spatial coding has two possible frames of reference, both of which are important to solving our everyday spatial tasks. Within each of these two frames of reference there are two available spatial systems, one simple and the other complex. The first frame of reference is viewer centered. In this format the individual is the focus and their task is about remembering how to do something—that is, which movements in space will achieve the goal. The simplest use of a viewer-centered spatial ability is called *response learning*. This is the system that is functioning when you reach for your coffee cup on the desk while you are writing. The cup is in the same place every day and you do not need to do more than execute a movement of the arm and hand to achieve the goal. Response learning is powerful and successful when the position of the viewer has not changed; however, it cannot account for movement. If my chair has been moved 6 inches to the right, my reach does not encounter the coffee cup anymore. A more complex use of the viewer-centered system is called *dead reckoning* or *inertial navigation*. In this system the location and direction of the viewer movement is tracked in spatial memory, providing a continuously updating sense of location and direction. This system can be best seen when we are navigating in the dark. The weakness in this system is that it is susceptible to small errors in calculation that are compounded over time. A misjudgment of distance or turn angle is updated and used in all subsequent estimations. This accumulation of error leads to larger and larger inaccuracies over time.

The other reference system is an external framework. There is a simple and complex way to use external information. The simple system, called *cue learning*, is the kind of navigation that allows us to use a landmark as a beacon to guide our search for the goal location. In this case, the landmark has to be at the desired location—for example, the keys are in the bowl. This system runs into difficulty when the landmark is not directly marking the desired location. In many cases, a number of landmarks as well as cues such as the shape of the environment provide a collection of pieces of information that when combined specify the desired location. This complex use of multiple sources of external information is called *place learning*.

These four systems have different starting points and developmental trajectories in humans. In general, infants were long thought to move from an egocentric (viewer centered) frame to an allocentric (external) frame. This view, first espoused by Jean Piaget, has the infant start out with response learning (if I look to my left when I hear door open I can see who is coming into my room), then moves to cue learning (if I look at the door when I hear the door open I can see who is coming). Cue learning is more accurate because it can account for any rolling around in the crib.

Research since Piaget has found this shift from the viewer-referenced systems to externally referenced systems to be less absolute than previously thought—that is, younger infants can show the externally referenced systems under supportive conditions far earlier than an absolute shift could be seen. These data indicate that the externally based spatial systems may be available much earlier than previously supposed. That they are not used under strenuous conditions may be related to the effort necessary to use them early in development or the weighting of the different systems across situations.

The adaptive combination model offers an account that can accommodate the finding of early capability. According to adaptive combination models, the four sources of spatial information are weighted such that the requirements of the situation can dictate which of the spatial abilities is most advantageous. In this way of looking at spatial cognition, the use of response learning to get coffee while typing requires minimal interruption of the concurrent task and thus is the best choice for the task at hand.

In development, the adaptive combination model allows for the idea that there may be some spatial abilities that are available but not used by the infant or young child because it is the weighting and reweighting of the spatial systems that is responsible for much of the developmental change through infancy and childhood. By this approach, infants have the ability to use several different spatial strategies from early on, and the task of infancy, instead of the shift from viewer-referenced to externally referenced systems, would be figuring out the
advantageous weighting of the available systems. The weighting of navigation systems changes dramatically when the child begins to locomote independently. The advent of crawling is associated with large cognitive change in several areas, including spatial cognition. The infant’s ability to move around space provides a new form of feedback, which changes the weighting of the spatial coding systems by providing new information about which system reliably gets the infant to a goal location.

In addition to the spatial systems available to the child and the weighting of those systems, there is also a developmental progression in the sophistication of the use and subdivision of otherwise unmarked spaces. Adults divide up unmarked spaces into a grid that allows them to search for a lost object in a smaller area. In effect, instead of remembering the car keys are on the football field, we divide the field into categories (between the goal line and the 10-yard line), which allows a smaller search area than if we did not divide the field. The hierarchical coding model indicates that estimations of spatial location are influenced by categorical spatial information as well as by a fine-grained estimation of location. Adults divide space into categories that, over a series of estimations, increase their accuracy by systematically biasing them toward the center of the category, thus allowing adult estimates of location to cluster in smaller subareas of a large space, instead of the whole football field. This bias works with the adult’s tendency to subdivide spaces into small categories to result in overall accuracy. This categorical bias increases when the fine-grained estimate is uncertain.

**Developmental Change in Spatial Ability**

Infants are sensitive to a large amount of spatial information very early in development. Infants as young as 3 months are sensitive to some spatial stimuli. They evidence longer looking times to objects that move from one spatial category to another. Five-month-old infants are sensitive to changes in object location but not to changes in shape or color of an object that remains stationary. At 9 months, infants show the A not B error. This occurs in an experimental paradigm in which the infant sees an object hidden in Location A and, usually after a short delay, is permitted to search Location A several times. Following several sequences of hiding at Location A, hiding the object at Location B in full view of the infant leads to another search at Location A. The A not B error has been intensively researched, and research indicates that the error can be seen earlier and later when delay between hiding at Location B and search is manipulated. At 16 months, children show a categorical bias in their estimates of location in a continuous space, particularly when their estimations are uncertain. At 21 months, infants show evidence of place learning, the most complex of the spatial systems.

Early childhood is a time of leaps in spatial ability. Young children become able to use basic spatial language early in the preschool years. Children also show that they are beginning to understand representations of spatial information, such as simple models and maps, and gradually they extend the number of situations in which they can use symbolic representations of spatial information. By the end of the preschool years, children show spatial perspective taking as long as they do not need to shift frames of reference. A child can now understand that the layout of the table will look different from someone else’s seat as long as they can maintain the same reference points.

Search patterns in children become more sophisticated through middle childhood. While 5-year-olds have difficulty with tasks that require them to use landmarks to accurately find a distal object (place learning) this ability improves through middle childhood. In a place learning task that allowed children to search a large-scale space presented on a desktop computer, 5-year-olds’ searches were very inefficient, generally relying on distance from the wall of the space, but the 6-year-olds used a single landmark. Neither of these strategies is very successful, but the addition of one landmark is a step toward the triangulation using multiple landmarks that is the hallmark of sophisticated place learning. As they move toward the end of middle childhood, children became more and more sophisticated in their landmark use with 8- and 9-year-olds using two distal landmarks to guide their search for an object that was distal from all the available landmarks. By the age of 12, the children use spatial strategies very similar to those seen in adults.

Map use also becomes more sophisticated during middle childhood, with children becoming able to understand mapping conventions and perspectives as well as dealing with more and more complex
maps. Middle childhood is when their ability to use spatial language moves beyond basic spatial terms to a more complex ability to account for the perspective of a listener and understand what a listener needs to know. By the end of middle childhood (10 years or so), children show mature hierarchical coding. They use the same categories as adults and their adjustment to the fine-grained estimates can occur along two dimensions at once.

Amy E. Learmonth

See also Representations, Development of; Visual Imagery; Visuospatial Reasoning

Further Readings


characterized thousands of human languages in terms of the vowels and consonants that each uses. Across all human languages, about 100 different such phonemes have been identified. Each language uses a subset of these, with some variation in the number across languages. English is a fairly typical language from this perspective, with about 42 different phonemes.

It has proven useful to think about each phoneme as being made up of a set of phonetic features. For example, it is possible to characterize each of the consonants in English in terms of three features: voicing, place of articulation, and manner of articulation. Voicing specifies whether or not the vocal cords are active during the consonant's production—they are active when producing a sound like /z/ but not when producing /s/. The place of articulation is based on where in the vocal tract the airflow is most constricted. For example, the air is completely stopped by the lips when saying /b/, whereas the restriction is in the middle of the mouth when saying /d/. Manner of articulation refers to how the air flow is restricted: For sounds like /b/ and /d/ (called stops) the air is fully stopped momentarily, whereas for sounds like /s/ and /z/ the air is only mostly restricted, with a bit slipping through; the noisy sound of the air escaping is called frication, and the manner is fricative.

The results of many experiments can most clearly be explained by considering the featural properties of the speech sounds. For example, if syllables are played to listeners under noisy conditions, and the listeners are asked to write down what they hear, the more features that two sounds share the more likely it is that one will be erroneously reported for the other.

**Factors That Make Speech Perception Difficult**

A number of factors complicate speech perception. One problem is a result of the way that consonants and vowels are produced. The position of the articulators for an intended vowel or consonant will be affected by the sounds that precede or follow that sound. This results in coarticulation—the acoustic properties of each vowel or consonant will not be the same each time, because the surrounding sounds are different.

The coarticulation problem is just one version of the general problem of phonetic variation: Each vowel, consonant, syllable, and word will potentially be different each time it is produced. Many factors influence such variation, including the particular individual who is speaking, the rate of speech, the semantic context, and so forth. Thus, there is no simple "template" that can be used that will consistently match the input.

An additional complication is that each word blends directly into the next word. That is, unlike this text, there are no "blank" spaces to separate words in speech. Any gaps in the speech stream are more likely to be due to producing a particular sound like a stop consonant than to breaks between words. This creates the segmentation problem—how does the listener know that a speaker said "two lips" rather than "tulips?"

**Possible Solutions to the Difficulties**

The coarticulation problem is quite fundamental, and some solutions to it are similarly fundamental. For example, in some theories listeners are not assumed to extract individual phonemes (which undergo the most extreme coarticulation). Instead, the basic recognition units are larger, such as combinations of consonants and vowels, or even whole syllables. These larger units are generally more constant in their form than the smaller pieces that make them up. There is also considerable evidence for compensation for coarticulation: If coarticulation shifts the acoustics of a phoneme in a particular direction, then listeners essentially shift perception back in the opposite direction, compensating for the coarticulation-induced changes.

The more general problem of phonetic variation seems to require a number of perceptual responses. The most general solution has been to use various normalization routines to make the (varying) input a better match to stored representations ("phonetic prototypes" or "perceptual magnets"). A second solution is to store all previously heard tokens of a given word. In such episodic models, a word can potentially be matched to hundreds or thousands of stored versions, increasing the chances of a successful match. Recently, a third approach has generated a good deal of research. Work on perceptual learning or recalibration has shown that, when listeners hear a variant of a speech sound that is far from the normal pronunciation, there is a tendency to expand the speech sound's category to encompass such
variation. For example, if a listener hears “alpha-
bet” with a /b/ that is acoustically rather like a /d/, then the /b/ category expands to include more /d/-
like sounds than before.

Listeners use several different methods to solve
the segmentation problem. For example, certain
sounds are produced differently at the beginning of
a word than in the middle of a word; hearing such
a variant provides an allophonic clue for the listener
to a word onset. In some languages, the stress pat-
tern of a word provides a clue for segmentation.
For example, in English, about 80% of the time, a
multisyllable word will begin with a stressed syllable
(e.g., donkey) rather than with an unstressed syllable
(e.g., delay). Listeners can also use word recognition
itself to help segmentation: Recognizing a word in
the speech stream provides segmentation informa-
tion (i.e., there is a word break just before and just
after this recognized word). Recent studies have
shown that under good listening conditions, this
lexically based segmentation strategy plays a major
role, whereas some of the lower level cues are more
important under noisier listening conditions.

Notable Speech Phenomena

Research on speech perception has revealed a num-
ber of interesting phenomena. When scientists first
began to study speech using relatively modern tech-
niques, they observed two apparently related phe-
nomena: categorical perception and the right-ear
advantage. Researchers created sets of syllables in
which a particular acoustic parameter was varied in
such a way that the syllable at one end of the con-
tinuum was heard in one way (e.g., /ba/), and the
syllable at the other end in a different way (e.g.,
/pa/). For simple nonspeech stimuli, varying a param-
eter this way leads to relatively continuous changes
in perception. For example, if one end of the con-
tinuum is a 100-hertz (Hz) tone and the other end is
a 200-Hz tone, with the intermediate items chang-
ing in frequency in a systematic way (e.g., 120 Hz,
140 Hz, 160 Hz, 180 Hz), listeners typically hear a
gradual change across the continuum; each tone is
a bit higher pitch than the one before it. For many
speech continua, in contrast, perception seemed cat-
egorical: Listeners heard a few items as one category
(e.g., /ba/), and then things abruptly changed, with
the remaining items heard as the other category (e.g.,
/pa/). This categorical tendency in perception was
strongest for stop consonants, somewhat weaker for
other consonants (e.g., fricatives), and weaker still
for vowels.

This same ordering was found in dichotic listen-
ing experiments, studies in which headphones were
used to play one speech sound to the right ear and
a different speech sound to the left ear. Listeners
showed a reporting advantage for speech played to
the right ear; as noted, the strength of this advantage
mirrored the ordering in categorical perception stud-
ies. Since the right ear has stronger connections to
the left hemisphere of the brain and language is gen-
erally processed on the left side, the right ear advan-
tage was taken as an index of specialized language
processing.

Another phenomenon that was discovered rela-
tively early was phonemic restoration. To produce
this effect, a small piece of speech (typically, one
phoneme) was cut out of a word, and a sound,
such as a cough or white noise, replaced the miss-
ing speech. Listeners consistently fail to notice that
speech is missing—they seem to perceptually restore
the missing speech. Similar effects have also been
reported for other complex sounds, such as music.
These effects suggest that the perceptual system can
use higher order information to help repair degraded
speech input, a valuable adaptation in the noisy
world in which speech must be heard.

When the speech input is ambiguous, listen-
ers have a strong bias to interpret it in a way that
yields real words rather than nonwords. Consider,
for example, a sound that has been designed to
be acoustically intermediate between /d/ and /l/. If
this ambiguous sound is followed by “ask,” listen-
ers generally report hearing “task” (a real word)
rather than “dask.” If the same sound is followed
by “ash,” listeners instead hear the sound as /d/ in
“dash.” This lexical bias is called the Ganong effect.

A final widely known speech phenomenon is the
McGurk effect, which is generated using audiovi-
ual presentation. The procedure involves showing
a headshot of someone producing a short utterance;
the audio track is dubbed to create a mismatch
between what the video shows and the sound that is
presented. For example, a video of the face produc-
ing /ga/ can be paired with an audio recording of /ba/.
Under these circumstances, listeners often hear /da/, a
kind of compromise between the visual and auditory
input streams. Many studies have explored how these
two sources of speech information get combined.
Stereopsis

Summary
Scientists have been studying speech perception for approximately 60 years. They have clarified the acoustic properties of the speech signal and identified several challenges that the signal potentially might pose for perception. Several solutions to these challenges have been identified. In addition, a number of interesting perceptual phenomena have been discovered, with these phenomena having the potential to constrain theories of how speech perception is accomplished.

Arthur G. Samuel

See also Hearing; Hearing, Philosophical Perspectives; Language Production, Incremental Processing in; Machine Speech Recognition; Word Recognition, Auditory

Further Readings

Stereopsis

As illustrated by the number and variety of topics covered in this encyclopedia, the human brain is a highly complex system. One of the true wonders is that it performs these feats simultaneously, and in most cases, without conscious effort. One such ability is that of stereoscopic depth perception. As described below, stereopsis is based on binocular disparity, one of many sources of depth information available in the environment. Other depth cues, such as perspective, shading, texture gradients, and occlusion, also provide clues as to the relative distances of objects in the environment, but none of these comes close to the quality and precision of depth percepts provided by stereopsis. The subsequent sections will review some of the defining features of stereopsis in humans and animals and its associated neural mechanisms.

Stereopsis is a cue to depth based on the fact that we have two eyes, which are laterally separated (in humans by about 6.5 cm). This positional difference in the two eyes results in each eye receiving a slightly different image of the world. So the image of one object will fall on slightly different or disparate retinal locations. This difference in location is referred to as binocular disparity and is the key information used by the stereoscopic system. Imagine that you have two cameras positioned side by side that take a picture of the same scene. While the resulting images will be very similar, there will be subtle differences, as illustrated in Figure 1.

In Figure 1 the observer fixates object F and another object M is positioned closer to the observer. If we trace the lines of sight from the objects to the back of the eye, we can see this positional difference (as illustrated at the bottom of the figure). Note that this binocular disparity information is generated in the same manner for objects in front of fixation (crossed) and beyond fixation (uncrossed). That is, the geometry of these arrangements is the same, but the position of the retinal images relative to the fixation point is reversed.

An important aspect of stereoscopic depth perception is that retinal disparity is generated between the object or point fixated and another object in the scene. Thus the perceived depth is relative to where one is looking, and if a third object is introduced at the same distance as the fixation point it will have zero disparity. In turn, zero binocular disparity indicates that an object lies on the plane of fixation. The set of points that are equidistant with a given fixation location define the horopter (e.g., the dashed line and point P in Figure 1).

The preceding description defines the theoretical horopter; however, when observers are asked
Stereopsis

To set stimuli to be equidistant with fixation, the shape of the horopter is not circular and instead is somewhat flattened. This is known as the empirical horopter. As retinal disparity is increased, there is a corresponding increase in perceived relative depth. Interestingly, over a large range of retinal disparities, the disparate images appear single, or fused. This region has been referred to as Panum’s fusional area. Beyond this range, the images are no longer fused but appear double or diplopic.

An important but often overlooked aspect of stereoscopic vision is that the upper limit for fusion, or Panum’s limit (the grey region in Figure 1), does not correspond to the upper limit for depth percepts from stereopsis. Instead, as documented by Armin Tschermak in the early 1900s, observers can reliably discriminate the relative depth of diplopic stimuli. It has been argued subsequently that fine (small) and coarse (large) retinal disparities are processed by different mechanisms. Current research supports such a distinction, though it remains possible that, while the quality of stereoscopic percepts changes at large disparities, the underlying mechanisms represent a continuum.

While binocular vision has been studied for centuries, since the time of Euclid (323–285 BCE), the link between retinal disparity and depth perception was not made until 1838. In a presentation to the Royal Society of London, Sir Charles Wheatstone revealed that the small differences in the views seen by the two eyes are directly responsible for three-dimensional (3-D) vision. Further he showed that it is easy to “deceive” the visual system and recreate the 3-D percept by separately presenting stereoscopic images to the two eyes (see Figure 2). This discovery initiated a new field of study in which investigators could measure the limits of stereopsis and the effect of a range of stimulus attributes on performance. Also, his invention led to the proliferation of handheld stereoscopes, which were used to view stereoscopic images of distant peoples and places. The current advancements in digital technology have led to a resurgence in popularity of stereoscopic entertainment in the form of 3-D movies.

**Stereoacuity**

Since Wheatstone’s discovery, our understanding of stereoscopic vision has grown exponentially. We now know that humans are able to discriminate the relative depth between two objects based on disparities as small as 30 arc sec (an arc sec [or second of arc] is one sixtieth of one degree). For highly trained observers this threshold can be as small as 4 to 8 arc sec. Such disparities are equivalent to (or smaller
Stereopsis than a hair's width separation between two needles viewed at arm's length! Psychophysical studies have shown that these low thresholds depend on a number of stimulus attributes. For instance, the smallest amount of disparity required to reliably discern that two points are separated in depth (stereoacuity) decreases with increasing image contrast and viewing time. Stereoacuity is degraded by factors such as blur and, in some instances, size.

The Neural Substrate of Stereopsis

In the 1960s, Horace Barlow and John Pettigrew were the first to identify binocular neurons tuned to particular retinal disparities, the possible neural substrate for stereopsis. Subsequent experiments in other laboratories showed that these disparity-selective neurons could be classified as tuned-excitatory, tuned-inhibitory, near, or far neurons. The excitatory and inhibitory cells responded to small, near-zero, retinal disparities, either with excitation or inhibition. The near and far cells preferred larger disparities and signaled only the direction of the depth offset relative to fixation. This discovery maps onto psychophysical results, which distinguish between the properties of disparity percepts in the small (fine) and large (coarse) range.

The critical link between the disparity-sensitive cells and stereoscopic-depth percepts was made by Randolph Blake and Joy Hirsch. They showed that kittens reared with one eye patched were stereoblind and had only monocular neurons in the primary visual cortex. Subsequent experiments of this type have shown that while the binocular neurons necessary for stereopsis are present at birth, concordant-binocular experience is necessary for the development of stereopsis. That is, the two eyes must move together so that the same region of the visual environment is seen at all times. This line of research has had important implications for the treatment of binocular disorders such as strabismus (misaligned eyes) and amblyopia (reduced vision in one eye).

More recent electrophysiological studies of stereopsis in primates have shown that the encoding of retinal disparity begins in area V1, but neural responses that correspond to perceived depth are found in area V2 and in higher cortical areas. Further, as one moves along the processing pathways, the complexity of the information encoded increases. One of the interesting aspects of ongoing research into the neural basis for stereopsis is the presence of disparity selectivity throughout many visual processing areas. It seems that there is no stereoscopic "center," and even areas believed to specialize in other image properties such as the motion-processing region—the medial temporal region (MT)—have a high percentage of disparity-selective units. Recent investigations of human stereopsis using brain scan technology such as functional magnetic resonance imaging (fMRI) have largely echoed the main results seen in the electrophysiology literature. One of the key topics in current electrophysiological and imaging studies is the possible division of stereoscopic processing into streams that map onto the dorsal/ventral categorization. In a recent study, Tim Preston and colleagues found that dorsal areas encode disparity magnitude, while the ventral stream encodes disparity sign or direction. This is an intriguing figure.

**Figure 2** Two pairs of stereoscopic images that can be fused to create 3-D percepts

*Note:* To view each pair of images, first cross the eyes to align the dark frames that surround the images. When they align in the center (you will see three images now), slowly focus your attention on the center image. It will, eventually, appear in depth. Once fused, the two lines in the upper panel will be displaced in depth with the top line behind fixation and the bottom line in front. The three ellipses in the lower panel will appear slanted in different directions in depth.
result that corroborates the proposed dissociation of coarse and fine disparity processing.

Conclusion
Since Wheatstone’s discovery of the link between retinal disparity and depth perception, much progress has been made in our understanding of human stereopsis, and, in particular, its neural substrate. Many questions remain, including the role of distinct dorsal and ventral processing streams in stereopsis, the extent of neural plasticity in disparity selective neurons, and how complex 3-D surfaces and structures are encoded. Answers to these and other open questions will require multidisciplinary approaches that combine rigorous psychophysical methods with brain scanning and electrophysiological techniques.

Laurie M. Wilcox and Debi Stransky

See also Depth Perception; Optic Flow; Perceptual Constancy

Further Readings

STROOP EFFECT

The Stroop effect (also called Stroop interference) is the phenomenon in which people are slow and error prone in naming the print colors of incompatible color words (e.g., when seeing the word yellow printed in red ink, people are to say “red”). Have you ever tried to remember the title of one song while another song is playing on the radio? You cannot seem to ignore the song on the radio, and so it interferes. Such interference is the bane of attention. Indeed, successfully attending in a world full of stimulation requires that we constantly ignore irrelevant stimuli so as to overcome interference with what is relevant. How might we study the phenomenon of interference, which so clearly influences our ability to process the world around us?

At the dawn of psychology, James McKeen Cattell documented that we are considerably slower to name objects or their properties than to read the corresponding words: Saying “table” to a picture or “yellow” to a color patch is slower than reading table or yellow aloud. Cattell saw this difference as evidence that word reading becomes automatic via extensive practice. Half a century later, John Ridley Stroop combined colors and words into a single task. When the task was to read the word aloud, ignoring the color, people had no difficulty compared to reading words in standard black ink. But when the task was to name the ink color aloud, ignoring the word, people had great difficulty compared to naming the colors of color patches.

In line with Cattell, word reading is taken to be so practiced that it has become automated, and hence words cannot be ignored—even when they should be. This indicates that we do not have absolute control over our attention: Attention can be attracted by the world (exogenous control), not just directed by oneself (endogenous control). The Stroop effect is the best-known evidence of this fact: It is one of the most robust phenomena in all of psychology and the basis of thousands of published studies.

What Causes Interference?
For 40 years, Stroop interference was explained as a kind of “horse race” with the wrong horse (the word) beating the right one (the color) to the stage where a response was prepared—a serial/sequential processing explanation. Thirty to 40 years ago, investigators began to suggest that interference results from performing a controlled process (color naming) simultaneously with an automatic process (word reading)—a parallel processing explanation. Then, about 20 years ago, with the advent of neural network (connectionist) models, emphasis shifted to the idea that learning occurred via changes in stimulus-response connection strength. In the last 10 years, theories have integrated the Stroop effect into larger scale models of perceptual processing or language processing, situating interference in broader cognitive perspective. These increasingly sophisticated models successfully encompass the many published results that constitute the empirical database for the Stroop effect.
Features of the Stroop Effect

After 75 years, we know a very considerable amount about the Stroop effect and, consequently, about the interference that arises when attention is not entirely successful—when ignoring fails. Critically, we know that interference is most likely to occur when there is disparity in practice on the two dimensions. Yet in studies where the color information has been presented sufficiently before the word to give the color response a head start, the Stroop effect does not “flip over,” such that the color begins to interfere with reading the word. So relative speed of processing each dimension (word and color) is not the whole story. But there are also empirical challenges to the automaticity explanation, such as the finding that introducing an additional word—not a color word—into the display reduces the interference. If reading is automatic, why should adding another word dilute the interference?

For decades, the Stroop task has been used as a hallmark index of attention; in fact, it is often part of neuropsychology test batteries. Six-year-olds are even familiar with it from the “Brain Age” series of handheld computer games! Moreover, there are many variations on the theme; for example, there is the picture-word task, where the object is to name a simple picture (e.g., a table) when an incompatible word (e.g., tree) is printed inside the picture. Interference can also be caused by noncolor words when they are activated—either acutely by recent encounter, or chronically by long-term exposure. This has led to numerous studies, most notably the emotional Stroop effect, where time to name print colors is greater for words related to an individual’s anxiety (e.g., crawly for a spider phobic; grade for a test-anxious person) than for neutral words (e.g., pencil), thought to result from chronic activation of the anxiety-related concepts and words. The emotional Stroop effect is used to diagnose anxiety disorders and even to measure the success of their treatment.

In the past 20 years, brain imaging techniques have been developed that provide information about the localization of cognitive activity, techniques such as functional magnetic resonance imaging (fMRI). When applied to the Stroop effect, such techniques have shown activity especially in the anterior cingulate cortex and the prefrontal cortex, areas now associated with cognitive control and, in the case of the Stroop task, the failure of that control. This control is viewed as implemented by the prefrontal cortex with the assistance of conflict monitoring done by the anterior cingulate cortex.

What is impressive is that such a superficially simple task, having been used for 75 years, still is useful in our exploration of how attention works. Attention is vulnerable to interference, and the Stroop task demonstrates this beautifully while also providing us with a way to understand that vulnerability better. With the ever-increasing research on the fundamental, cognitive mechanisms of attention, especially on their neural underpinnings, this venerable task will continue to be studied and may even become a more important tool in our cognitive toolkit.

Colin M. MacLeod

See also Attention, Neuroimaging Studies of; Attention and Action; Attention and Consciousness; Attention and Emotion; Automaticity

Further Readings


Subliminal Perception

This entry first provides a definition and overview of the issue of subliminal perception. It then gives a brief history of this issue followed by a discussion of the scientific debate surrounding it. Finally, more
recent trends related to subliminal perception are presented.

**Definition and Overview**

The term *subliminal perception* refers to one’s ability to perceive stimulation below the *limen*. The limen refers to the amount of intensity at which the stimulus can be noticed half the time. That is, our sensory systems are not capable of detecting all the stimulation present in the environment. The stimulation has to reach some intensity before it can be noticed. Subliminal stimulation refers to situations in which a stimulus is presented at an intensity below the limen; hence, the stimulus is seldom, if ever, perceived with awareness.

The *American Heritage Dictionary* provides two definitions for *perceive*. The first is “to become aware of directly through any of the senses, especially sight or hearing.” The second is “to achieve understanding of; apprehend.” The implication is that when sensory information is perceived, one may become aware of the stimulus provoking perception (i.e., supraliminal perception), or one may gain some understanding of the stimulus even in the absence of awareness (i.e., subliminal perception). This latter possibility has intrigued researchers for over a century: Is it possible for one to perceive information that somehow alters our understanding of the world without our awareness of said perception occurring?

**Brief History**

This question of whether stimuli presented in a subliminal manner can be perceived despite the absence of awareness was the focus of the first published article from a psychology laboratory in North America. In 1884, Charles Pierce and Joseph Jastrow asked participants to guess whether cards contained letters or digits after first establishing a presentation distance wherein participants claimed to be unaware of what was on the cards. Guessing performance was above chance, which the authors attributed to some entity other than the “primary waking self,” perception in the absence of awareness.

This issue became of interest to the general public in 1957 when an unemployed market researcher named James Vicary claimed to have subliminally presented the words “Drink Coca Cola,” and “Eat popcorn,” during a movie, resulting in an increase of product sales. Later, Vicary admitted his study was a marketing stunt with the intent of increasing the number of movie goers. However, the report spawned the worry that it might be possible to influence human behavior without their awareness, a worry reinforced by subsequent books such as Wilson Bryan Key’s *Subliminal Seduction; Ad Media’s Manipulation of a Not So Innocent America.*

**The Scientific Debate**

Given the public interest, psychological researchers began attempts to scientifically document, or refute, the existence of subliminal perception. Some argued there was clear evidence for subliminal perception, but critics countered that the methodologies employed were insufficiently rigorous to sustain such claims. The crux of these disagreements centered on the procedure that was used to document subliminal perception.

To demonstrate perception of a stimulus presented at energy levels low enough to preclude awareness, one must first devise a scientific way of measuring awareness and then find an energy level at which this measure indicates null sensitivity. Once established, if some other measure indicates the stimulus is being processed, then subliminal perception has been demonstrated. This methodological approach is termed the *dissociation paradigm*, as the goal is to dissociate some general measure of perception from a more specific measure of perception leading to awareness. The study by Pierce and Jastrow highlighted previously provides such an example in the sense that participants’ claims were used to indicate awareness, and their guessing performance provided the more general measure of perception. When guessing performance remained above chance despite claimed unawareness, the general measure of perception was dissociated from the specific measure of perception resulting in awareness.

The controversy with respect to subliminal perception centers primarily on the extent to which researchers accept certain definitions and measurements of awareness. If one allows the participant to indicate when they are and are not aware of some stimulus (a “subjective” index of awareness), then it is relatively easy to demonstrate the perception of a stimulus that participants claim to be unaware of it. However, if one insists on an “objective” index of awareness and further insists that this measure
Synesthesia

Synesthesia is a neurological condition in which stimulation of one sensory modality or cognitive pathway leads to automatic, conscious experiences in a second, unstimulated pathway. For example, in music → color synesthesia, auditory inputs cause synesthetes to see colors, which typically include movement and texture. In one of the most extensively studied forms of synesthesia, grapheme → color synesthesia, letters and numbers are experienced with a colored overlay. Synesthesia research has expanded dramatically in the past 20 years. This research has demonstrated that synesthesia is a real phenomenon, explored its neural basis, and begun to uncover the genetic mechanisms that might lead to synesthesia. Studies of synesthesia are relevant not only to understanding how individual differences in neural structure lead to unique perceptual experiences but also to understanding universal processes.
of cross-modal integration and the philosophical riddle of qualia, or the raw subjective feels of sensory experience.

Although synesthesia was a topic of intensive scientific, artistic, and cultural interest in the late 1800s and early 1900s, it was largely forgotten until the late 20th century when research into individual differences and subjective internal experiences once again became widespread. Renewed scientific interest in synesthesia also arose as new behavioral and neuroimaging methods demonstrated the reality of synesthetic experiences. In the past 20 years, there have been more published studies on synesthesia than in the entire preceding century of research.

**Behavioral Studies**

Behavioral studies have demonstrated that synesthetic associations within an individual are present from childhood and are stable over long periods of time, with synesthetes being greater than 90% consistent in the associations they report, even after years, compared to 30% to 40% consistency after just a month in nonsynesthetes, even when participants are warned they will be retested. Interference paradigms, including modified versions of the Stroop effect, demonstrate that synesthesia occurs automatically. For example, when a grapheme → color synesthete who experiences the digit 5 as green is presented a 5 in red ink, she is slower to name the ink color than if it were printed in green ink. Similar paradigms have been used to demonstrate the automaticity of numerous other forms of synesthesia.

Visual search and segregation paradigms have demonstrated that synesthetic colors can improve performance for synesthetes. In one early test, Vilayanur Ramachandran and Edward Hubbard presented synesthetes and nonsynesthetes with a matrix of 5s in which a number of 2s were embedded to form a hidden shape: a square, diamond, rectangle, or triangle. For a synesthete for whom 2s are red and 5s are green, for example, the display appears as a red triangle on a green background, which improves synesthetes’ ability to identify which shape was embedded in the display. Numerous other studies have demonstrated that while synesthesia is elicited early in perceptual processing, it does not occur prior to attention. Additionally, there are substantial individual differences in the intensity of the colors experienced by different synesthetes.

**Neural Mechanisms**

The neural mechanisms of synesthesia are still debated. Some researchers have suggested that synesthesia arises as a result of decreased synaptic pruning between adjacent brain regions, while others argue that synesthesia is a result of decreased cortical inhibition. Functional neuroimaging studies (including positron emission tomography and functional magnetic resonance imaging) have demonstrated increased activation in color-selective areas including V4 when both music → color and grapheme → color synesthetes are presented with synesthetic triggers. Additionally, neuroimaging studies using methods that focus on brain structure, including diffusion-tensor imaging and voxel-based morphometry, have demonstrated anatomical differences in brain regions involved in eliciting different forms of synesthesia. Grapheme → color synesthetes show increased connectivity in regions of the inferior temporal lobe associated with visual processing, while a unique synesthete who experiences tastes in response to different musical intervals (e.g., she reported that a major third was sweet, while a minor sixth tasted of cream) showed increased connectivity in brain regions associated with auditory and taste processing. These studies showing anatomical differences in synesthetes’ brains are consistent with the pruning hypothesis but could also arise because of plastic changes as a result of decreased inhibition. Multiple, neural mechanisms may be involved in synesthesia, and different forms may depend on different mechanisms.

**Prevalence and Familiality**

Recent estimates of the prevalence of synesthesia suggest that it may be as common as one in 23 people across all its forms. Early studies suggested that synesthesia was more common in women than in men, leading to the suggestion that synesthesia might be inherited through an X-linked mechanism. However, subsequent random sampling has demonstrated that synesthesia occurs equally commonly in men and women, arguing against the X-linked mode of inheritance. Although Francis Galton recognized that synesthesia runs in families in the 1880s, the first candidate genes for synesthesia have only recently been identified. Future research will be required to confirm these findings and to better understand their roles in brain development.
Broader Implications
Although different synesthetes report different associations, large-scale studies have identified trends in synesthetes’ experiences. For example, the letter A is more likely to be associated with red, smaller numbers and higher pitches are associated with brighter colors, and words tend to share tastes with the phonemes that make up the food names. Such systematic mappings are also found in nonsynesthetes but do not reach conscious awareness, suggesting that some of the same mechanisms underlie synesthetic and nonsynesthetic associations. In addition to its inherent interest as a perceptual variant, synesthesia is of interest to philosophers because it may shed light on the question of qualia: Synesthetes experience additional qualia evoked through nonstandard pathways. The implications of the existence of synesthesia are still debated by philosophers.

Edward Michael Hubbard

See also Music Perception; Stroop Effect; Word Recognition, Visual

Further Readings

SYNACTIC PRODUCTION, AGREEMENT IN

In linguistics, agreement refers to the correspondence of some formal feature (person, gender, number) between an agreement controller (e.g., the subject noun) and syntactically related words in the sentence (e.g., the verb). Experimental psycholinguistics capitalized on “attraction” errors resulting from interference of an intervening element in agreement production to shed light on the internal dynamics of syntactic constraints in language production. For example, in Sentence 1 the verb are erroneously agrees with the plural intervening noun cabinets.

1. *The key-S to the cabinets-P are-P on the table.
(S = Singular, P = Plural, * = ungrammatical sentence)

This entry summarizes the major findings about the structural conditions determining interference effects in agreement and their theoretical implications for models of syntactic production.

Interference by Syntactic Features
The first major observation is that, by and large, only syntactic features of the intervening element (like its number or gender features) have the potential to trigger interference. Features on the noun may also be represented conceptually (e.g., plurality) and morphophonologically (i.e., in the word form, like the final s on most plural nouns in English). However, it was found that the representation of agreement features at these levels, when manipulated on the intervening noun, play no role in interference (although they were found to influence agreement when manipulated on the agreement controller). For example, conceptually plural but syntactically singular interveners (e.g., The coach of the team . . .) fail to trigger plural agreement on the following verb. Similarly, no interference arises with syntactically singular pseudo-plurals that carry typical morphemes of plurality (like the final phoneme /z/, e.g., the color of the rose). In contrast, the recurrent finding that syntactically plural interveners (e.g., the daughter of the neighbors) generate stronger interference than singular interveners (e.g., the daughters of the neighbor) was used to support the claim that plural nouns possess a syntactic feature (marked) capable of triggering interference, whereas singular nouns lack any such feature (default).

Interference Within Hierarchical Structures
The second major empirical finding is that interference occurs within the hierarchical structure and not on the surface word order. The hierarchical structure of the sentence reflects its internal organization: Words combine into phrasal units (e.g., the noun phrase combines the determiner and the noun),
which themselves combine to form higher order constituent units (e.g., the prepositional phrase combines the preposition and the noun phrase). The treelike representation is illustrated in Figure 1.

In the marking and morphing model of agreement developed by Kathleen Eberhard and colleagues, interference arises as features from the intervening element are incorrectly transmitted onto the agreement target during the morphing stage. The process of feature transmission is assumed to operate at a stage of language production where elements are organized hierarchically. A first line of evidence comes from the cross-linguistic observation that elements situated higher in the hierarchy intervene more than those situated lower. For example, when two elements intervene in the surface order between the agreement controller and target, the element situated higher interferes with agreement (programs in “The computer with the programs for the experiment are broken”), not the element situated lower (experiments in “The computer with the program for the experiments is broken”), although the latter is linearly closer to the verb. Interference was also found to be stronger with prepositional phrase modifiers (e.g., “The editor of the history books”) than with relative clauses (e.g., “The editor who rejected the books”), the former being situated hierarchically higher than the latter. Nevertheless, when hierarchical height is kept constant as in disjunctive constructions (e.g., “The boy or the girls”), interference is strongest with the element linearly closer to the verb, showing that linear proximity may also play a role.

**Fine-Grained Syntactic Modulation of Interference**

Evidence for the role of more fine-grained aspects of the hierarchical structure in interference comes from the observation that the interference occurs in the absence of any intervention in agreement in the surface word order. For example, interference was reported with prepositional phrase modifiers in interrogative sentences involving verb movement (e.g., “Are-P the helicopter-S for the flights-P safe?”) and with objects in cleft or relative constructions involving object movement in French (e.g., “John speaks to the patients-P who(m) the medicine-S cure-P”). In such sentences, the object has moved from its postverbal position to a frontal position, which does not intervene between the subject and the verb on the surface. These effects contrast with the absence of interference found in constructions that have exactly the same surface orders but different underlying hierarchical structures (free
inversion in Italian declarative sentences, e.g., literally, “Phones-S the friend-S of the neighbors-P,” and complement clauses in French, e.g., literally, “John tells the patients-P that the medicine-S cures-S”). Critically, these latter constructions, in contrast to interrogative or relative clauses, fail to involve syntactic movement. The finding that interference effects occur specifically in structures with moved objects was argued to provide evidence for the role, in language production, of the construct of syntactic movement developed in linguistic theory.

In sum, experimental research on interference in agreement strongly argues for a model of syntactic production in which operations like agreement are realized under the guidance of syntactic features over structurally defined hierarchical configurations.

Julie Franck

See also Language Production, Agreement in; Planning in Language Production; Production of Language

Further Readings


Taste, Philosophical Perspectives

Taste has been unjustly neglected in the philosophy of perception, largely as the result of the failure to recognize the complexity of tasting experiences. This complexity has been brought to light through research by sensory scientists, which offers philosophers the opportunity to reexamine traditional thinking. This entry will consider traditional views of tastes and tasting, review key empirical findings, and examine the consequences of such research for philosophical views of tastes as subjective or objective.

Tastes as Sensations or Properties of Substances?

Taste has received little attention in the philosophical study of the senses, partly because it has traditionally been considered one of the lower bodily senses, doing little more than producing sensations in us when we eat and drink. On this view, taste is not a perceptual sense like vision or audition, which gives us information about the external environment; it is an inner sense giving us information only about ourselves and our subjective responses to foods or liquids. Although we taste something when we bite into an apple or sip a cup of coffee, such tastes are thought to be no more than pleasant or unpleasant experiences, something immediately known on the basis of sensations on the tongue, and these sensations are often thought to be too fleeting, too variable, and too subjective to be revelatory of anything beyond themselves.

However, we need not equate tastes with individual responses in the taster. A more objectivist view would see tastes as properties foods or liquids have and that we are able to perceive by tasting. This view seems closer to common sense than the traditional philosophical picture, for we appear to rely on taste to give us knowledge of the flavors of things we eat and drink, to tell us whether a strawberry is ripe and whether our coffee has sugar in it, and to distinguish between the taste of an apple and the taste of an onion. It is hard to imagine how we could come to have this knowledge by any other means.

Those who defend the subjectivity of taste think of tastes as sensations in us whereas defenders of the objectivity of tastes think of the tastes as properties of a food or liquid. How we adjudicate between subjectivist and objectivist positions depends on how we characterize tastes and tasting. In the traditional picture, the experience of tasting is relatively simple, amounting to no more than having sensations on the tongue in response to the items we consume. However, this view is questionable in its supposition that a tasting experience is due to the workings of a single sense. The case for objectivity requires a more detailed account of our tasting experiences. Such an account is to be found in cognitive psychology and neuroscience where, far from being peripheral to our understanding of the senses, the experience of tasting is thought to provide key insights into the nature of perception.
The Psychology and Neuroscience of Tasting

A growing body of evidence indicates that the senses do not operate in isolation but typically interact to produce integrated multisensory experiences. Tasting is no exception, and contrary to what we commonly assume, the experiences we have when tasting foods are not produced by the single sense of taste alone but by the integration of information from touch, taste, smell, and perhaps other sense modalities. Touch gives us information about the texture of what we consume. We can describe a sauce as creamy or a wine as viscous. The remainder of what we ordinarily call taste is actually a mixture of taste and smell, though subjects are not able to consciously separate these components in their experience. Smell provides the largest part of what we call taste, which is why people who lose their sense of smell often think they cannot taste anything. When questioned, patients will admit they can taste basic gustatory qualities like salt, sweet, sour, bitter, savory, and metallic, but everything else that is missing from their taste experiences is due to smell.

The coming together of information from different sensory streams produces the unified experiences we have when eating a peach or drinking wine. So although we are right to think of tasting as giving us a distinctive kind of sensory experience, we fail to recognize such experiences as involving several sensory components, of which taste is just one. These multimodal experiences of the qualities of foods and liquids have become the focus of intense scientific research, providing as they do clues to the understanding of the cross-modal influence of one sense on another and the multisensory integration of inputs from different senses. The integration of taste, touch, and smell is known as flavor to distinguish it from the purely gustatory taste component, which we seldom, if ever, experience in isolation. Flavor describes the sapid and odorous properties of a substance, including its temperature and texture, as well as its power to irritate the trigeminal nerve—the nerve whose activation is responsible for our finding chili “hot” and menthol “cool” despite there being no change of heat in the mouth. So when speaking about the taste of a food, we are actually speaking about its flavor. This point is often missed because we fail to notice the components of our tasting experiences and because we are unaware of the large role smell plays in sustaining them.

Taste and Retronasal Olfaction

It is easy to dispute the claim that what we taste depends on smell when focusing on orthonasal olfaction, the external part of smell where we inhale odors and gain information about the environment. The internal part of smell is retronasal olfaction, experienced as we exhale, giving us information about what we have just eaten on the basis of odors traveling from the mouth or gullet through the nasopharynx to receptors in the olfactory epithelium at the top of the nose. The odors we sense orthonasally by inhaling may be experienced quite differently when they are sensed retronasally as we exhale. Pungent cheeses, for example, can be much more palatable in the mouth than would have been predicted when smelling them orthonasally. The integration of taste and retronasal olfaction along with touch produces the characteristic experiences of flavor. If we block retronasal olfaction by preventing odors in the mouth from reaching receptors in the nose, blindfolded subjects cannot distinguish the taste of an apple from that of a raw potato.

Taste and Orthonasal Olfaction

The ability to distinguish the components that contribute to an experience of something’s flavor depends, to some extent, on the taster’s powers of discrimination. However, some smell and taste components of flavor are experientially inseparable. The fusing (or confusing) of smell and taste can be demonstrated by purely odor-induced tastes, where subjects experience a tasteless liquid as sweet or a sweetened liquid as sweeter when it is accompanied by an aroma of vanilla presented orthonasally. Such odor-induced sweetness can even suppress the sourness of a liquid as real sweetness does, leading some sensory scientists to consider such aromas as tastes. This “sweetness enhancement” effect is highly robust and persists even when subjects are asked to attend to each sensory component separately. The inability to distinguish sensations of taste and retronasally sensed aromas is understandable, but in this case subjects fail to discriminate between tastes and orthonasally sensed aromas presented simultaneously.

Cross-Modal Effects in Tasting

Tasting provides several examples of the cross-modal effects of one sense on another. Certain retronasal
aromas can make substances taste creamier. The colors liquids have can influence the perception of flavor. The high-frequency sound of our own crunching, when boosted, can make stale potato chips “taste” fresher. Low temperatures can accentuate bitterness as we notice when coffee goes cold. The greater a liquid’s viscosity the greater its perceived sweetness. Should we treat such cross-modal effects as illusions or just routine aspects of human tasting experiences? (Notice that cross-modal effects leave a trace in our language, such as when we describe vanilla as “sweet-smelling” even though sweet is detected only by taste receptors and vanilla is not itself sweet.)

Aftertastes provide further evidence of the conflation of taste and smell. We experience aftertastes as being in the mouth even though they are retro-nasally sensed odors. Purely olfactory stimuli can be experienced as tastes and not smells, as is easily demonstrated by putting a tasteless aqueous jelly with olfactory properties into the mouth, which will cause subjects to report that they are having sensations of taste on the tongue even when the experimental conditions are known to them. Such cases illustrate what is known as the location illusion, where we relocate or refer olfactory sensations to the mouth. The retronasal detection of odors, when accompanied by sensations of touch from the tongue, leads to the referral of sensations of smell to the mouth where they are interpreted as tastes.

The Temporal Dimension of Tasting

A further dimension to flavor experience is its dynamic time course. Tasting is not a simple occurrence but an unfolding process with a separate series of stages. Different flavors are detected at different places in the mouth, thus tasting experience’s dynamic time course affects what we can pick out when. Sensory characteristics will change across time depending on where they are experienced. This complex sequence allows us to build up a profile of the food we consume by attaching different hedonic responses to the different aspects discerned and to the experience as a whole. It is an activity by which we assess the things we eat and drink. Expert wine tasters, for example, pay particular attention to what happens at each stage, which gives them clues about the qualities of a wine. Novices, by contrast, taste in a different way and are unlikely to be aware of different stages, thereby missing much of the detail. In this way, how we taste affects what we taste. When assessing food and drink there is room for individual variation, not only because people differ in what they like and dislike but because they may have different experiences as a result of the different thresholds they have for the basic taste qualities of sweet, salt, sour, bitter, savory, and metallic. So-called supertasters have heightened sensitivity to some of these qualities and frequently find unpleasant what the rest of us enjoy.

The Subjectivity and Objectivity of Tasting

How does this newly revealed complexity in our tasting experiences bear on the issue of whether tastes are subjective or objective? At first, it appears to put pressure on the subjectivist view of taste. The subjectivist supposes that tastes are personal experiences had by individuals in response to what they ingest or imbibe and immediately knowable on the basis of gustatory sensations alone. The initial appeal of this view is due to the failure to recognize the underlying complexity of our tasting experiences. But how do matters stand for the objectivist about taste? The objectivist can point out that there is more to what is call taste than we notice at first. We can miss or be in error about features of our experience, thus showing that in the domain of tasting, how things appear to us is not always how they are. This gap between appearance and reality opens up space for a more objectivist view of tastes and the experience of tasting.

However, the subjectivist can reply that the gap just exposed is between experience and how we think about it, not between our experience of tasting and what it is experience of. The reply is telling for both parties. On the one hand, the concession about the often unnoticed gap between our experience and the immediate assumptions we make about it seems to undermine the subjectivist’s unproblematic entitlement to the materials she relies on to make out her case. On the other hand, the objectivist needs to do more to show that foods and wines genuinely possess the flavors that we are better able to recognize by improving our perceptual acuity.

The objectivist can point out that since overall flavor perception depends on touch, itself a perceptual sense, flavor perception must be a perceptual sense. Subjectivists may reply that while the products of
touch accompany the products of taste (and smell), they are not part of what we call taste. This reply would be unconvincing since tactile experiences do seem to make an essential contribution to our tasting foods. We describe 2-day-old potato chips as tasting stale, when the only difference in flavor between the fresh and the stale crisps is a difference in texture. Texture appears to play a constitutive role in flavor perception.

Is Flavor a Psychological Construct?
But what of the flavors perceived? Are they really aspects of the external environment or just psychological constructs? The latter view amounts to a scientifically informed version of subjectivism about taste. Flavors could be an amalgam of sensations produced by different sense modalities united into a single percept. Such a view is still wedded to the idea that all we recognize are properties of our own sensory experiences. However, the texture properties of foods we describe as creamy, crunchy, or viscous are not properties of our sensations but properties of the foods in our mouths.

The objectivist about taste can point out that flavors are not traditional secondary qualities, like sounds or colors, which can only be detected by one sense. It takes at least three senses to pick out flavors. This does not make them common sensibles, like shape, detectable by more than one sense, since none of the contributing senses detects flavors on its own.

A frequent objection to objectivism about taste is the individual variation in subjects’ judgments about foods and wines. However, care is needed in pressing this objection. First, the complexity of tasting experiences provides the objectivist with reasons to expect, and the resources to explain, the variety of reactions. Second, talk of variation is often exaggerated. Milk that has soured tastes disgusting to all, and a banana could not taste like an orange to some people without there being something wrong with such people or with the banana. Disagreements about how things taste usually concern complex flavors. Third, we must distinguish between how something tastes and whether it is to someone’s taste. Individual preferences may vary a lot but it does not follow that the particular flavors people like or dislike vary similarly. Of course, we cannot rule out that the wine you like and that I dislike tastes differently to you and to me. But this may be due to a number of factors, including how skilled we are as tasters and whether we have vastly different thresholds for certain compounds. Thus, variation may be due to something other than that individuals simply have different responses to the same tastes.

Flavor Perception: One Sense or Many?
Why suppose that the unified experience of tasting misleads us about its multisensory nature? It could be argued that what we call taste is a single sense and that the complexity revealed by neuroscience simply concerns neural mechanisms that subserve given sensory modalities. This view would incorporate retro-nasal olfaction into taste, even though it makes use of the same olfactory receptors as smell, thus dividing it from orthonasal olfaction, which would now be taken to exhaust the sense of smell. This strategy faces many problems. Not least, it fails to capture generalizations about the interaction between taste and smell. Why, for instance, do individuals with a poor sense of smell have a poor sense of taste? And why do people who lose their sense of smell report losing their ability to taste? Finally, what explanation can be given of the sweetness enhancement effects?

A very different approach is to ask whether there is a single flavor sense, over and above its component senses. This approach treats the sense of flavor as a perceptual system that guides successful food selection by picking out flavors as multidimensional properties of things in our environment.

Conclusion
Tasting may yet prove one of the most illuminating experiences for philosophers of perception to work on, revealing as it does a hidden complexity to our experience. Work in this area will require knowledge of the empirical findings of sensory scientists and an account of how information from separate sensory streams are integrated into a single, unified percept. It is an important area of interdisciplinary research and more work is needed.

Barry C. Smith

See also Smell, Philosophical Perspectives

Further Readings
Teleology

Teleology is the study of purposes, goals, or ends; a teleological explanation explains a process or behavior by stating the goal toward which it was directed. This is especially relevant to the study of mind, because part of what it is to be a creature with a mind is to act for purposes, goals, or reasons. To act for reasons is to exhibit goal-directed behavior; it is to have one’s behavior be teleologically explicable. This entry explores the role of teleology in the explanation of human behavior, contrasting teleological and causal accounts of action explanation and closing with a discussion of attempts to reduce teleological explanation to other forms of explanation.

Explaining Human Action

Typical teleological explanations take forms like the following:

- $A$ did $B$ in order to $G$.
- $A$ did $B$ for the sake of $G$.
- $A$ did $B$ for the purpose of achieving $G$.
- $A$ did $B$ to $G$.

So, for example, we might say that Kristen went to the kitchen in order to get a beer; this means that the agent (Kristen) directed her behavior (her going to the kitchen) toward the state of affairs in which she has a beer.

When we explain the behavior of persons by citing their reasons, our explanations are often in explicitly teleological form. However, we also sometimes give reason-based explanations where the explanations are not in this form. For example, if we say, “Kristen went to the kitchen because she wanted a beer,” then it appears that we have explained the behavior by citing an antecedently existing mental state (her desire for a beer) rather than by citing some goal state toward which her behavior was directed.

The Causalist View

There is an active dispute among philosophers of mind concerning the nature of these different forms of explanation of human action, and the dispute can serve as the dividing wedge between very different views of mind and agency. On the one side are the causalists, who maintain that explanation of action in terms of reasons is a species of causal explanation. Thus, when we say that Kristen went to the kitchen because she wanted a beer, we are saying that Kristen’s desire for beer caused her behavior of going to the kitchen. The causal account of action explanation then becomes the cornerstone of broadly reductionist accounts of mind, according to which facts about the mind are ultimately reducible to physical facts.

The causalist position is generally dominant within contemporary philosophy of mind, but it is less obvious than it might appear. For starters, even if we say that the agent went to the kitchen because she wanted beer, this does not by itself imply that the explanation is causal. Other uses of the word because are clearly not causal—for example, “the argument does not work because it equivocates on the key term.” In this example, the point is that we are justified in saying that the argument does not work on the basis of the fact that there is an equivocation; we are not claiming that the equivocation is a cause of invalidity in the way that germs are a cause of disease. So analogously, from the fact that we might say that Kristen went to the kitchen because she wanted a beer, we should not automatically conclude that her desire for beer is being cited as a cause of her behavior, at least not if cause is being used in the normal sense associated with the physical sciences.

Moreover, other commonsense explanations of behavior are in explicitly teleological form; for example, “Kristen went to the kitchen in order to get a beer,” and such explanations do not even cite an antecedent mental state at all. The causalist must say that such teleological explanations are nonetheless best construed as or reduce to causal explanations. However, there have been persistent problems in getting any such causalist reduction to work, and it is by no means clear that the problems here are merely technical. This will be further discussed below.
The Teleological View

Opposing the causalists are those who take teleological explanations of action to be basic and irreducible to other forms of explanation. In the teleological view, goal direction is an ineliminable phenomenon. On this view, explanations such as, "Kristen went to the kitchen because she wanted a beer," are construed as teleological explanations. The reference to her desire for beer serves to specify the goal—namely, that of getting a beer in the way she desires. We need not deny that Kristen's behavior had a cause, and it seems exceedingly probable that the causal chain leading to her behavior crucially involved various states of her brain. However, on the teleological account, our commonsense, reason-based explanation is not aiming to identify the cause of the behavior. The teleological explanation is simply answering a different question: not “What was the antecedent cause of the behavior?” but “Toward what end was Kristen directing her behavior?” Both of these questions might be put in the words, “Why did Kristen go to the kitchen?” but the questions are nonetheless distinct.

Moreover, these distinct questions likewise arguably involve distinct methods of inquiry. When answering the teleological question about an agent's purpose in acting, we take a broadly interpretive approach. We attempt to make as much sense of the person as we can. We try for a theory of the agent on which she is, broadly speaking, as rational as possible, meaning that she believes what she ought to believe and values what she ought to value. Accordingly, ascertainning the goal toward which the behavior was directed, we seek candidate explanations on which two things are true: First, the hypothesized goal is such that the agent's behavior is appropriate for achieving that goal, and second, the hypothesized goal is of comprehensible value for the agent. When making these judgments, we naturally take into account facts about the agent's circumstances and epistemic situation. For example, it might be that there is no beer to be found in the kitchen, and thus Kristen's behavior of walking to the kitchen is bound to fail at the supposed goal of getting a beer. However, her behavior can still be appropriate for that goal if Kristen reasonably believed that there was beer in the kitchen.

In routine cases, it is often quite obvious which goal to cite in teleological explanation. If Kristen has just been told that there is beer in the kitchen, and she says, “Oh, good, just what I want,” then it is clear that her subsequent behavior of going to the kitchen would be appropriate for obtaining a beer, and it is clear enough that having a beer would have value from Kristen's perspective. Of course, things might not be so simple. If an annoying relative just entered the room, then Kristen's real goal might have been to avoid the relative, and the beer could be just an excuse. Or perhaps her behavior was genuinely directed toward both states of affairs—getting a beer and getting away from the relative. Further data about her behavior, including what she says and thinks to herself, will help us in making the most rational sense of the behavior.

On this approach to teleology, there is an irreducibly normative element that makes teleological explanation quite different from causal explanation. When investigating the motions of a rock or of a planet, we are not constrained by the normative requirement that we make the planet or rock's behavior be as rational as possible. Or to put it the other way around, if we were to try to make rational sense of a planet's behavior, we would fail. We might say that the planet had the goal of following the laws of physics, and then its behaviors would be appropriate for the goal, but it would be hard to see why following the laws of physics would have value for a planet or anything else. In a teleological account, being an agent requires a complex set of goals or a life. We cannot successfully attribute anything of the sort to the planet. So we conclude that the planet is not an agent at all, and no teleological explanation of its behavior will be true.

Reductive Accounts of Teleology

The teleological account of action explanation is quite contentious. Some philosophers claim that teleological explanations can be reduced to causal explanations. For example, one might suggest that A did B in order to G is true if and only if

$A$ had a desire for $G$ and a belief that by Bing she could $G$, and this belief and desire caused $A$'s Ging.

However, this analysis appears to be inadequate. To borrow an example from Alfred Mele, a nervous philosopher at a conference desires to distract her commentator and believes that she could distract the commentator by knocking over the pitcher of water
on the table. The very fact that she finds herself with this desire and belief unnerves her to the point that her hand shakes uncontrollably and her shaking hand knocks over the pitcher of water. Here, it seems that the belief and desire play the causal role required in the analysis above, but we would not in fact conclude that she actually knocked over the water in order to distract her commentator; the behavior was involuntary and not directed at anything. There can and have been ingenious attempts to patch up the causal analysis, but these attempts themselves seem to run into similar problems.

Other philosophers try to reduce teleological explanations, not in a straightforward causal way but by noting their similarity to evolutionary or selectional explanations. Many biological explanations appear to be in teleological form:

Birds have wings in order to fly.

Pandas have a “thumb” in order to strip the leaves off of bamboo.

In cases like this, the form of explanation is species $K$ has trait $T$ for purpose $G$. Behind such an explanation lies an evolutionary story: Creatures of species $K$ had ancestors who developed trait $T$ through mutation, and this trait allowed those ancestors to accomplish $G$, which in turn led to higher differential reproduction by members of the species with trait $T$. Thus $T$ came to dominate the population. One can put this by saying that $K$s have $T$ in order to $G$, but the teleological form is clearly shorthand for the evolutionary story. Clearly, there need be no agent involved, and this is not a case of irreducible teleology. One might then try to see teleological explanation of human action along similar reductive lines.

There will, however, be substantial obstacles to this sort of reduction of teleology. First, the explanatory pattern above concerns the relative frequency of traits, and the reductionist would need to apply this model in some way to individual items of behavior. Presumably, one can give evolutionary explanations for dispositions to behave certain ways in certain circumstances—for example a cat’s disposition to arch its back in the presence of perceived threats. However, there will be obstacles to a complete reductive account of rational behavior along these lines. The basic problem is that all evolutionary explanations come back to items being naturally selected because of their tendency to lead to greater differential reproduction of genes, whereas it is not the case that all rational goals that are of conceivable value come down to this one aim. Rational animals can do all sorts of perfectly reasonable things that do not benefit the reproduction of our genes. Indeed, some seemingly quite rational actions are inimical to that end—for example, using birth control.

The debate concerning teleology in mind is still quite active, and there is no consensus on whether teleological explanations can be reduced to more naturalistic causal or selectional explanations. If the nonreductionists are right, then within the realm of creatures with rational minds, purpose is an ineliminable feature of the world, and facts about mind will not reduce to physical facts.

Scott Sehon

See also Action and Bodily Movement; Explanation of Action; Mental Causation; Philosophy of Action

Further Readings


Theory of Appearing

When we open our eyes or employ any of our other senses, physical objects in the world appear to us. The theory of appearing is a theory of what it is
for a physical object to appear to a conscious subject. Preliminarily, we may say that when a physical object appears to a conscious subject, it is related to that subject in a certain way, so we can speak here of the appearing relation. The theory of appearing holds that the appearing relation is a unique relation, a relation fundamentally different from all other kinds of relations. Consequently, it is committed to the view that the appearing relation is not a causal relation. The theory of appearing is controversial because one might think that the appearing relation has to be a causal relation. This entry describes and motivates both the theory of appearing and the opposing theory that the appearing relation is a causal relation, the causal theory of appearing.

The Causal Theory of Appearing

One might think that there are only a limited number of ways in which distinct entities can be related to each other: They can stand in spatial relations to each other, temporal relations, and/or causal relations. Consider now the appearing relation. Suppose I open my eyes and see a tomato; in virtue of seeing the tomato, the tomato visually appears to me. A tomato cannot appear to me unless it exists at the same time as myself and is situated within my field of vision. But surely to say that the tomato appears to me is to say more than that the tomato is temporally and spatially related to me in a certain appropriate way. So the appearing relation must be at least in part a causal relation. According to the causal theory of appearing, what it is for a physical object to appear to a conscious subject is for the physical object to cause the conscious subject to undergo a certain kind of conscious state: a perceptual experience. Consider again the tomato that is appearing to me. Note that the tomato does not merely appear to me in some general manner; it appears to me in virtue of appearing to me in particular ways: It appears red to me, for example. According to the causal theory of appearing, the tomato appearing red to me is a matter of the tomato causing me to have a certain kind of visual experience, a visual experience somehow characterized by redness. Science gives us many details about the nature of this causal process (the tomato reflects light into my eyes, the light stimulates receptor neurons in my retinas, etc.), but we need not be concerned with these details here. What is relevant here is that according to the causal theory of appearing, all there is to a physical object appearing to a conscious subject is the object causing the subject to have a certain kind of experience. The experience itself is merely a state of the subject; the subject is related to the physical object only in that the physical object causes the subject to have the experience.

The Theory of Appearing: Response

The advocate of the theory of appearing opposes the causal theory of appearing because she opposes the idea that experiences are merely states of conscious subjects. According to the theory of appearing, these experiences are themselves relations between physical objects and conscious subjects, and it is these experiential relations that constitute the appearance relations. On this view, there are relations other than spatial, temporal, and causal relations. What motivates the advocate of the theory of appearing is the commonsense idea that our experiences encompass the physical objects we perceive. Here I am, looking at the tomato that appears red to me and having a visual experience. As noted earlier, the visual experience is characterized in part by redness. Part of what is involved in my having this visual experience is my being aware in some sense of this redness. But this redness, although an element of my experience, also strikes me as being a feature of the tomato itself; the tomato itself seems to be included in my experience (this is the commonsense idea to which I previously referred). The tomato is appearing to me in virtue of the fact that one of its features is present in my consciousness. Specifically, the tomato is appearing to me in virtue of appearing red to me, and it is appearing red to me in virtue of its redness being present in my consciousness. I am related to the tomato via its redness; the redness of which I am aware is itself a relation between the tomato and myself, and it is this relation that constitutes the appearance relation between the tomato and myself. Some will find it strange to speak of redness as a relation; surely there are no color relations in the same sense as there are spatial relations. But note that a physical object cannot appear red without appearing red to a subject; for a physical object to appear red is for it to be related in a certain way to a subject. And the theory of appearing holds that all there is to a physical object appearing red is its presenting its redness in a subject's consciousness, so this redness itself must be
of such a nature as to relate the physical object to a conscious subject.

We can now characterize the theory of appearing more generally as the view that what it is for a physical object to appear to a conscious subject is for the physical object to present one or more of its sensory features in the subject’s consciousness. The challenge for the theory of appearing is to show that it is compatible with a scientifically informed picture of the world. This challenge is taken up in the readings listed below.

Harold Langsam

See also Disjunctive Theory of Perception

Further Readings


**THINKING**

Thinking refers to the process of reasoning in order to reach a goal. In humans, this process typically involves combining externally derived information and prior knowledge so as to formulate and evaluate implications that may provide an answer to a question or a solution to a problem. It is the goal-directed nature of thinking that sets it apart from mere associative processing, where one idea links to another in a nonpurposive manner akin to what takes place when daydreaming. Thinking is a core topic of empirical inquiry and theoretical analysis in cognitive science and subsumes a multitude of interrelated concepts, including reasoning, categorization, judgment, decision making, hypothesis testing, problem solving, and creativity. Of all these interconnected concepts, however, reasoning is arguably most central to understanding what thinking entails.

This entry begins by summarizing key historical antecedents to research on thinking and reasoning and then progresses to consider important theoretical insights deriving from contemporary research in this field. These insights are discussed with reference to a major paradigm that has been deployed over several decades in researching thinking processes: the four-card selection task developed in the 1960s by Peter Wason. The entry concludes by considering some important trends in current thinking research.

**Historical Antecedents to Contemporary Thinking Research**

The study of thinking extends back over 2,000 years to Aristotle, who believed that it was the conscious activity of the mind, with thoughts being composed of images. Aristotle also pioneered the method of *introspection* to study thinking, a technique that was dominant in philosophy and psychology until the late 19th century. Aristotle’s view that images are the foundation of thinking was central to the associationist accounts of the British Empiricist School of philosophy in the 17th and 18th centuries. This view only became discredited when psychologists at the University of Wurzburg in the early 20th century demonstrated that image-based thoughts did not characterize the thinking of many participants, with some describing no discernible thoughts at all and others claiming their thoughts were indescribable and seemingly nonconscious.

Research in the 20th century further undermined the notion that thinking relates to conscious processing. Freudian theory advanced the idea of unconscious thinking as an essential determinant of behavior, while behaviorists such as J. B. Watson and B. F. Skinner contended that all behavior, including thought, could be explained in terms of individuals learning to associate particular responses with particular stimuli when a reward was present that reinforced such links. From a behaviorist perspective, analyzing the conscious, “mentalistic” correlates of thinking was an irrelevance, with thinking instead being described as reflecting acquired habits and conditioned responses operating at a tacit level.

In the 1960s, the field of cognitive psychology emerged, with its basis in a new computational metaphor for the mind and a resurgence of interest in the mental processes underpinning thinking—an interest that continues unabated. Although the cognitive revolution meant that the study of thinking was back on the agenda as a legitimate area of inquiry, this approach made no commitment to the view that thinking is necessarily conscious and
available for introspective access. Indeed, there has long been recognition amongst cognitive psychologists that implicit processes may dominate thinking, with only surface features emerging in the stream of consciousness. The cognitive perspective on thinking additionally avoids limiting such activity to humans, such that certain machines (e.g., artificial intelligence systems) can be viewed as engaging in thinking, as can certain animal species (e.g., higher order primates). Cognitive researchers have also tended to avoid treating human thinking as synonymous with notions of rationality, given abundant evidence that thinking often appears to be irrational and suboptimal. Finally, the cognitive approach brought with it a renewed interest in the mental representations underpinning thinking. Although the concept of images has featured in cognitive theorizing, a rather different concept has burgeoned over the past 30 years, which is the idea espoused by Philip Johnson-Laird that thinking is based on the construction and manipulation of abstract “mental models” of possible situations.

**Thinking: An Example Paradigm and Findings**

The previous definition of thinking describes it as involving goal-directed reasoning. Reasoning, or *inference*, has a long tradition in philosophy in the fields of logic and probability and emphasizes the process of drawing implications or conclusions from given information (premises). A valid *deductive* inference is one that produces a conclusion that must be true given the truth of its premises. Deduction is closely related to formal logic, which provides a *normative* model against which deductive thought can be assessed. While deductive reasoning is truth preserving, *inductive* reasoning is not, instead providing only plausible conclusions that may or may not be true. The strength of induction resides in its capacity to enable the formulation of conjectures that go beyond the available information, allowing, for example, the generation of generalizations or laws based on repeated observations of events.

**The Wason Selection Task**

The four-card selection task developed by Peter Wason in 1966 is certainly the most investigated paradigm in the history of thinking research, perhaps because it has all the hallmarks of a useful task for studying thinking, including a stated goal and a need for hypothesis testing, deductive inference, and decision making. In its standard, abstract form (Figure 1) the task involves presenting participants with four cards that are described as each having a letter on one side and a number on the other side. The presented cards display the facing sides A, J, 3, and 7. Participants are also given a conditional sentence, “If there is an A on one side of a card, then there is a 3 on the other side,” and are asked to decide which card or cards need to be turned over to determine the truth or falsity of the sentence. Common choices are A or A and 3. The logically correct choice (which few participants make) is A and 7, since only a card with an A on one side that does not have a 3 on the other side would disprove the sentence; hence, selecting the A and the 7 is necessary to reveal such a potentially falsifying combination.

Pioneering research by Jonathan Evans established that responses on this task primarily reflect a “matching bias,” a tendency to select cards named in the presented sentence. This was corroborated using sentences involving negated terms (e.g., “If there is not an A on one side of a card, then there is not a 3 on the other side”). Although negations change the logic of the task, thereby altering the cards that should be selected, participants still tend to select A and 3, in line with matching. These selection task results are curious since they suggest that human thinking may be rather superficial in nature, showing limited sensitivity to logical principles of sound inference. Evans proposes that matching bias dominates our “intuitive” reasoning on the selection task by directing attention in a highly selective way toward aspects of the presented information. However, Evans and Linden Ball also present evidence
indicating that more conscious, analytic processes are still engaged on the task but primarily function to enable people to find good reasons to justify the selection of cards cued through intuitive processes. The influence of matching bias on the selection task is so powerful that few participants (typically less than 10% of undergraduates) are able to overcome it so as to choose the logically correct cards. Those individuals who do choose correctly have superior intelligence and an apparent ability to override intuitive processing by the application of what Evans refers to as the “reflective mind”—that is, the capacity to think in an abstract and hypothetical manner that is not merely dominated by the specific content and context of the task at hand.

Keith Stanovich has presented his own concept of the reflective mind, which is different in important ways from the notion discussed by Evans. Stanovich views the reflective mind as being the “disposition” to engage in explicit, analytic reasoning—that is, a person’s preference for careful, analytic deliberation over quick, intuitive judgment. Furthermore, he views the disposition to engage in analytic thinking as being distinct from the actual “capacity” to execute analytic thinking in an effective manner, what he refers to as the “algorithmic mind.” Presumably then, the high level of intelligence needed to reason logically on the abstract selection task is a manifestation of the joint activity of both the reflective and algorithmic minds.

The selection task can be made much easier if it is recast in a real-world format. One variant (Figure 2) involves giving people a social rule, “If a person is drinking beer then he or she must be over 18 years of age,” and presenting cards representing four drinkers. One side of each card depicts what the person is drinking; the other side depicts that person’s age. The presented cards have facing sides showing Beer, Coke, 22 years of age and 16 years of age. Participants have to decide which card or cards need to be turned over to discover whether the rule has been violated. The majority correctly chooses the person drinking beer and the person under the age of 18. Intriguingly, success on this version has little association with intellectual ability. Evans argues that on the standard selection task the intuitive mind is prompting the wrong answer and hindering the efforts of the reflective mind to apply logical reasoning. In contrast, on the realistic problem, the intuitive mind is cueing the correct answer, taking the pressure off the reflective mind. Indeed, most people will have previously encountered something similar to the drinking-age rule and will know from experience that rule breakers are those who drink alcohol when underage. As such, little thought is required to select the correct cards, which is why individuals of higher intelligence have no advantage over those with lower ability on this version.

### Current Trends in Thinking Research

Research on thinking and reasoning has produced considerable evidence for so-called dual-process theories, such as the one sketched out above in relation to the abstract selection task, where implicit, intuitive processes interact with explicit, reflective processes in controlling responding. Evans describes his own dual-process theory as the “two minds hypothesis.” He conjectures that the intuitive mind is old in evolutionary terms, sharing features with animal cognition, whereas the reflective mind is recently evolved and distinctly human. The intuitive mind is also claimed to be the source of emotions and intuitions, capturing adaptive behaviors acquired over evolutionary history as well as habits acquired experientially. In contrast, the reflective mind enables abstract thinking so as to facilitate reasoning about hypothetical possibilities.

Critically, dual-process theorists propose that intuitive and reflective processes will frequently come into conflict, and when this happens it is the intuitive mind that often wins out, with the reflective mind seemingly rationalizing the conflict such that people appear to be unaware of the fact that their intuitions are dominating their thinking. This phenomenon is not only seen in the abstract selection task but in many other thinking and reasoning
paradigms where intuitive/reflective conflicts can arise. One particularly good example concerns the study of “belief bias” in deductive reasoning, where people frequently make intuitive judgments about presented arguments in accord with the believability status of given conclusions rather than making reflective, analytic judgments in accord with the underlying logic of the arguments. A recent demonstration of this comes from a study by Edward Stupple and colleagues. They showed that the incorrect tendency for people to endorse believable conclusions to invalid arguments is primarily associated with individuals of moderate or low analytic ability, who engage in less reflective thought (as indicated by relatively rapid response latencies) compared to those of high analytic ability, who take more time over their reasoning in an attempt to resolve validity/believability conflicts.

Dual-process theories of thinking are currently widespread, with key areas of investigation relating to understanding the complex interplay between intuitive and reflective processes in determining how we reason when working toward goals. Dual-process notions are not, however, universally accepted, with some theorists arguing for a more unitary view of thinking processes. Especially dominant in this vein is the Bayesian rationality approach of Mike Oaksford and Nick Chater, which proposes that thinking involves implicit, probabilistic calculations that have no relation to deductive logic. Such research has been gaining popularity in its attempt to provide a comprehensive account of all human thinking, including the dominant pattern of card choices on both standard and realistic selection tasks. It is difficult, however, to see how probabilistic theorizing can be reconciled with data indicating that participants often do make an effort at deductive thinking. Another dominant trend in thinking research relates to the examination of the neural underpinnings of reasoning using brain-imaging techniques. Interestingly, current neuroscientific evidence supports the existence of competition between distinct brain systems during many aspects of thinking, thereby further corroborating dual-process ideas.

Linden J. Ball

See also Deductive Reasoning; Two System Models of Reasoning

Further Readings

Time Perception

Time perception refers to the subjective experience of the duration or temporal organization of events within a given period of time. In this entry, a brief overview of time perception is provided, including the main phenomena and models used to interpret them. The main neurobiological substrates of timing identified in recent studies are then presented.

Various types of temporal experience can be distinguished: the main types being perception of duration of events or stimuli, perception of order (which between a and b came first or second), and perception of temporal regularity or rhythm. These phenomena can take place on different time scales, varying from a few milliseconds to seconds, hours, days, and even years. For example, musical rhythm is perceived with a series of tones separated by temporal intervals shorter than about 2 seconds, and natural rhythms are perceived in succession of days, months, or years. Perceiving duration, order, and
temporal regularity is fundamental in most activities for organisms evolving in a changing environment. Classical Pavlovian conditioning experiments as well as later experiments on animal timing demonstrated that pigeons, rats, and many other nonhuman species adapt remarkably to temporal contingencies of the environment, displaying highly developed abilities in interval timing and revealing a key role of temporal associations in learning. Things that are judged to be close in time tend to be associated, and this link constitutes the basis of learning in humans and other animals. Estimating time and processing temporal order are also essential in performing complex activities requiring coordination in movements or action, anticipation of times of occurrence in a dynamic environment, or remembering ordered elements like digits in phone numbers or words in sentences. In addition to being intimately related to most common activities involving planning, coordination, and memory, the ability to estimate time is also essential in rhythmic activities such as playing music or dancing.

A stopwatch is an efficient mechanism for estimating duration: It starts and ends at distinct moments, and the amount of temporal information accumulated between these two points constitutes an objective assessment of the interval bounded by the two points. Humans can quite accurately perform that kind of interval timing without any external time-keeping device. Interval timing is flexible in that it can start and stop any time in response to the demands of the environment, in contrast to rhythmic timing, such as in circadian rhythms, which is often determined by rather rigid constraints and shows relatively small variability. The phenomena, methods, and models described below mostly concern human interval timing in the few hundred milliseconds to minutes range.

**Interval Timing in Humans:**

**The Main Phenomena**

Our subjective experience of time does not correspond necessarily to objective time, as measured by an accurate clock. The first experimental studies on the relationship between perceived and objective time were performed in the 19th century by psychophysicists such as Gustav Fechner, Wilhelm Wundt, and Ernst Weber. One issue considered important by these scientists was whether time perception shared common features with perception of other dimensions, such as visual perception or auditory perception. Even though time perception cannot be related to a specific sensory system like visual or auditory perception, some principles indeed seem to apply to time perception as well as to perception of visual or auditory features of stimuli. One major principle is *Weber's law*, which states that the *just noticeable difference* (jnd) between two stimulus values (e.g., line length, light brightness, or tone duration) is a constant proportion of the smaller of the two values. It will be easier to notice the 1-second (s) difference between 1 s and 2 s than the same difference between 50 s and 51 s. Although the jnd appeared to be an increasing function of the smaller duration values, in more recent studies on time perception the relationship is not exactly linear over all values (especially when shorter than .25 s or longer than 2 s) and is better described by a generalized form of Weber's law.

Differences between judgments of very short and longer intervals (e.g., shorter than .25 s and longer than 2 s) suggest that different mechanisms may be responsible for estimating intervals of various ranges. Perception of short time intervals is influenced by their sensory content. One example is the filled-duration illusion: A filled interval is usually perceived as longer than the same empty interval. For example, a .25 s tone will be judged longer than a silent .25 s interval between two brief markers. There are also some differences in judgments of short stimuli of different modalities. One difference is that people judge an auditory stimulus to be longer than a visual stimulus of the same duration. However, although judgments of short durations are influenced by the sensory systems used in the timing task, most current models of time perception assume that a common mechanism underlies time perception in the various modalities.

Judgments of longer durations—that is, longer than about 1 s—require storing temporal information relative to the ongoing duration in working memory. Attention is another cognitive process involved in judgments of longer durations. The influence of attention is especially obvious in interference studies, which reveal that when people are estimating the duration of a time interval, performing some attention-demanding task at the same time
perturbs considerably the precision of their estimate. The interference usually results in a shortening of the perceived duration, and the longer or the more attention-demanding the concurrent task, the shorter the perceived duration. A related finding is that the mere expectation of a signal interferes with timing. As with concurrent tasks, the longer the duration of expectancy, the shorter the perceived duration. The interference effect is generally attributed to the fact that precise timing requires continuous attention, so if a distracting activity is performed while timing, then some temporal information is lost, resulting in shorter perceived duration.

**Methods and Paradigms**

The main methods used in human time perception studies are duration comparison or discrimination, production or reproduction of time intervals, and verbal estimation. Duration discrimination involves posing a relative judgment—for example, deciding which is the longer of two tones. In time production or reproduction, a participant may be asked to reproduce the duration of a tone by pressing a key for a duration equal to the presented duration. Most methods can be used in two general paradigms, retrospective and prospective. In the retrospective paradigm, people estimate in retrospect the duration of a past temporal interval; this paradigm may be required if during the stimulus presentation they were not aware that an estimate of its duration would have to be provided. In the prospective paradigm, people estimate the duration of a stimulus during its presentation. Different factors seem to affect performance in the two paradigms: Retrospective judgments are more dependent on the memory of the number of events or perceived changes taking place during the elapsed duration; prospective judgments are strongly influenced by the amount of attention devoted to time during the ongoing duration. This difference corresponds to the paradoxical observation that a past period of time may seem long in retrospect when many things happened during that period but that time flies when we are busy or having fun. Different models thus attempt to explain experimental data obtained in the retrospective and prospective paradigms, respectively.

**Models of Prospective Timing**

Prospective temporal estimates in humans are approximately related to objective clock time. In “internal-clock models,” on which this section mostly focuses, this relationship is explained by assuming the existence of a mechanism similar to an interval timer in humans and other animals. Like a clock, this mechanism would emit temporal units (often named pulses) that are related to real time in an orderly way. The *scalar expectancy theory* (SET) is an influential internal-clock model of time estimation that includes three categories of processes related respectively to the clock, memory, and decision. Although initially developed from experiments with rats and pigeons, this model appears to be very useful in interpreting results of experiments with humans. In this model, the clock level includes a pacemaker emitting pulses that are stored in an accumulator if a switch is closed, permitting accumulation. The content of the accumulator is transferred to working memory if the task requires a memory representation of the current time. The current time value is compared with a remembered criterion time sampled from a reference memory in order to evaluate whether the ratio between the current and remembered values is small enough to decide that they correspond. The decision rule varies according to the particular task requirements. Some experiments in animal timing suggest that the switch permitting accumulation of temporal information may be attention controlled, which would also explain that distracting activities shorten perceived time in humans by preventing accumulation of temporal information. Attention is also central in other internal-clock models of human time estimation, in which the amount of temporal information accumulated is positively related to the amount of attention devoted to temporal processing.

In the first internal-clock models, pulses emitted by the pacemaker were assumed to be the basis of temporal information in humans, but it has been suggested more recently that oscillatory processes constitute this basis. Periodicities related to oscillatory processes—for example, electrical cortical oscillations—are present in the bodies of humans and other animals. These processes present regular phases and would provide the basic information of the internal clock.

Humans and other animals may also use their own behavior to estimate time (finger tapping in humans, wheel running before pressing a lever for rats). Some behavioral models of timing assume that behavior supports time estimation
Two System Models of Reasoning

Abundant evidence suggests that there are two distinct systems of human reasoning, which can be referred to as intuitions and deliberations. The evidence comes from studies of deductive and inductive reasoning, decision making, categorization, problem solving, probability and moral judgment, and planning. Most of the evidence is behavioral, although a little comes from cognitive neuroscience. The distinction rests on a set of properties that characterize each system (see Table 1). The intuitive system is designed to make quick and dirty assessments based on similarity and what can be directly retrieved from memory. It relies more on observable properties and well-ingrained schematic knowledge. The deliberative system is slower and more analytic. It depends directly on learned systems of rules, and its information processing is highly selective. We have conscious access not only to its products but also to its inner workings. The intuitive system is likely more evolutionarily primitive than the deliberative system, has more in common with other animals, and includes a greater proportion of older brain structures. This

Further Readings


Neurobiological Substrates of Timing

Many brain areas have been identified in studies on the neurobiological substrates of timing; the specific areas often depend on the timing tasks and the range of durations investigated. One challenging issue, especially in studies on timing of intervals longer than a few hundred milliseconds, is to identify brain regions activated specifically by timing functions, independent of associated task demands such as memory or decision making. When humans are asked to perform perceptual and motor timing tasks, the main structures identified in functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and electrophysiological studies include the basal ganglia, prefrontal cortex, cerebellum, supplementary motor area, and some other discrete cortical areas that vary with the specific timing task requirements (stimulus modality, decisional, and response demands). In research with patient populations, perception and production of series of brief time intervals were disturbed in patients with cerebellar lesions, suggesting the involvement of the cerebellum in time estimation. Finally, pharmacological and lesion experiments as well as data from patients such as people with Parkinson’s disease also suggest a central role of dopaminergic activity in timing. Drugs facilitating and blocking the synaptic release of dopamine alter temporal performance in such a way that they seem to speed up and slow down the clock, respectively.

Claudette Fortin

See also Attention, Resource Models; Music Perception; Reinforcement Learning, Psychological Perspectives; Working Memory
entry provides an overview of the evidence for this characterization.

Characterizing the Systems

**Deductive Reasoning**

The distinction between intuition and deliberation helps characterize how people think in almost every area of cognition that has been studied. To illustrate, deductive inferences such as determining what follows from “if p then q” and “p is true,” can be made either way. Deliberation leads to more correct judgments of logical validity, but correct inferences require more processing time and more attention than intuitive inferences. They are thus less likely in the face of attention-demanding secondary tasks. Even without distractions, people are biased when judging the validity of arguments in favor of conclusions they believe to be true; their intuitive beliefs inhibit their ability to analyze whether a conclusion follows logically from an argument’s premises. People are sensitive to instructions; for instance, requests to respond deductively versus inductively change which system dominates, but people do not seem able to rely exclusively on deliberation while ignoring their intuitions.

Studies using functional imaging demonstrate that different brain areas become activated depending on whether a task demands associative responses or rule use. Some studies of deductive inference have suggested that a left temporal pathway corresponds to one reasoning system while a bilateral parietal pathway underlies the other. But other researchers have compared probabilistic reasoning using a task that involves both intuition and deliberation with a deductive reasoning task that relies more heavily on deliberation and found that both activate the medial frontal region bilaterally as well as the cerebellum. Probabilistic reasoning activated the left dorsolateral frontal regions more and deductive reasoning activated right occipital and parietal regions more.

**Decision Making**

People differ in which system they habitually use to make decisions. Some people are more likely than others to inhibit incorrect intuitive responses in order to make more deliberative decisions. Such people tend to make choices that map more closely onto the expected value of options. In gambles that promise gains, they are more risk seeking than less deliberate people, and in gambles that promise losses, they are more risk averse. But deliberative reasoning can sometimes lead to worse decisions. Because of the limited capacity of working memory, deliberation is only able to consider a few attributes of each option. Therefore, intuition is better equipped to make decisions when there are many relevant attributes and is better at accommodating attributes that are difficult to verbalize or quantify. Some believe that intuition is closely related to affect, although little evidence supports this claim.

**Categorization**

Categorization can be either rule based or similarity based. Rule-based categorization classifies based on no more than a few dimensions, is easily applied to novel stimuli, and is learned and used explicitly. Similarity-based categorization aggregates over many dimensions, generalizes only with a measure of uncertainty, and cannot be verbalized. People are more likely to use rule-based processes when they need to explain or justify their responses.

Using different methodologies, several brain-imaging studies have found distinct activation patterns for the two types of categorization, though the activated regions corresponding to rule- and similarity-based processing have varied. A study using artificial visual stimuli found activation in the medial temporal lobe for rule-based category learning and in the basal ganglia for similarity-based category learning. Another study used meaningful category
labels and found increased activation in frontal areas for rule-based categorization but not for similarity judgments. A study using novel animal-like stimuli showed increased activation in left inferior frontal cortex and anterior cingulate when explicit rules rather than similarity to prototypes were used for classification.

**Problem Solving**

People solve problems in two ways, using intuition (the “a-ha” experience) or deliberately and analytically. People can predict how long a problem will take to solve but only when it lends itself to analytic and not intuitive processing. Some types of problems, such as math problems, are inherently symbolic and appropriate for rule-based analytic processing. Greater working memory capacity helps solve such problems. Sian Beilock and her colleagues have shown that pressure to perform can impede performance, especially for individuals with more working memory resources.

**Judgment**

Several phenomena of judgment provide evidence for and help to characterize dual systems of thinking. The most prominent example is the conjunction fallacy of probability judgment, the observation that an event (e.g., that a flood will devastate Manhattan sometime this century) is sometimes judged less probable than a conjunction that includes that event (e.g., that global warming will cause glaciers to melt and a flood will devastate Manhattan sometime this century) even though this contradicts the prescriptions of probability theory. This exemplifies *simultaneous contradictory belief*, a phenomenon in which we are predisposed to believe that the conjunction is more likely even after we have discovered that it is not logically possible. People with lower IQs are more likely to commit the conjunction fallacy as are people engaged in a secondary task, again suggesting that working memory capacity must be available to engage in deliberation.

Moral judgments also have at least two bases, a deliberative one and a more intuitive one. People are only able to justify some of their moral judgments (e.g., committing a bad act is worse than not acting even if the outcome is the same). Other justifications cannot be articulated. For example, moral judgments tend to conform to the principle that harm intended as a means is worse than harm foreseen as a side effect, but people are not able to articulate that principle. This suggests that some moral judgments are made on the basis of intuition, although people have access to a system sometimes capable of generating justifications. Joshua Greene and his colleagues have given people a variety of moral dilemmas while imaging their brain activation using fMRI. Some dilemmas were solved based on simple rules, others based on feelings that people could not justify. The former were associated with areas of the brain usually associated with working memory, the latter with areas usually associated with emotion and social cognition.

**Planning**

In research on planning and action, the two systems are evident in the difference between novel and routine action plans. Sometimes we break goals down into subgoals, a deliberate process that is verbally accessible and working memory intensive. In contrast, actions that lead to a familiar goal, such as brushing your teeth, involve acting out learned schema. This can happen with very little awareness or cognitive effort while engaged in a parallel task. Such intuitive planning can result in errors. A stimulus in the environment can evoke a familiar action but one that does not achieve the current goal (e.g., eating another cashew when you intended to stop). Such action slips are common in certain frontal lobe patients and occur in healthy individuals when working memory is occupied by another task—for example, when deliberation is unable to intervene.

**Factors Governing System Choice**

The deliberative system is more resource intensive than the intuitive system, so intuition dominates when resources are scarce. Increases in time pressure or working memory demands increase the proportion of intuitive responses. Decreasing physiological resources such as sugar has the same effect. After an energy-depleting task, people who drank lemonade with sugar made more rule-based decisions than those who drank lemonade with an artificial sweetener.

Mood also influences which system guides behavior. When people are sad, they deliberate more and
Two System Models of Reasoning

make more consistent decisions. When people are happy, they integrate more general knowledge into their thinking, are more risk-averse when choosing gambles, and make better intuitive judgments, all suggesting greater intuitive processing.

Work by Danny Oppenheimer and his colleagues has shown that disfluency, the sense that a task is difficult, increases the degree of deliberative processing on that task. Disfluency is a metacognitive signal that a difficult problem requires deliberation. However, it increases deliberative processing even when incidental—for instance, when a reasoning problem is printed in a blurry font or participants are asked to furrow their brows (suggesting cognitive effort).

Conclusion

The distinction between deliberative and intuitive thinking has helped explain phenomena from a vast assortment of cognitive domains. All these explanations refer to a deliberative system that requires effort and working memory, which is selective in the information it uses but can abstract away from specific content, and that provides conscious access to the process in addition to the result of a computation. These explanations also refer to an intuitive system that is automatic and unhampered by concurrent tasks, that integrates information associated with specific content, and that provides only the result of a computation to conscious awareness. Some theorists argue that rather than two distinct systems, reasoning varies on a continuum between intuitive and analytic processes. Others argue that the many distinctions made between two forms of reasoning do not map onto the same two systems.

Magda Osman and Ruth Stavy show that children use some rules without deliberation or effort in ways that conflict with other rules. The rule-based system is not constituted by a consistent logical system. Rather, different rules emerge depending on how problems are framed, what the focus of the question is, and what is most available to the problem solver. Nevertheless, the evidence suggests that there is a single deliberative system. Whether there is a single intuitive system or a class of loosely related intuitive processes is not so clear.

Imaging data suggests that distinct brain mechanisms can be engaged by varying instructions, but the evidence does not clearly support any specific theory of the neuroanatomy underlying deliberation or intuition.

One open question concerns the relation between intuition and emotion. Another concerns how the systems interact. Jonathan Evans has distinguished several models of their interaction. Most of the evidence favors the parallel-competitive model, which proposes that the intuitive and deliberative systems process information in parallel and any conflict is resolved after both processes generate potential responses. But there is also reason to believe the default-interventionist model, which states that the intuitive system always operates first, followed by an optional deliberative intervention and override or elaboration of the intuitive response. Of course, both models might be partially correct because cognitive processing can involve many cycles of interaction.

Steven Sloman and Adam Darlow

See also Automaticity; Categorization, Neural Basis; Decision Making, Neural Underpinnings; Deductive Reasoning; Mental Effort; Representativeness Heuristic; Thinking; Working Memory

Further Readings

UNCONSCIOUS EMOTIONS, PSYCHOLOGICAL PERSPECTIVES

How do you feel right now? Do you feel happy, sad, disgusted, fearful, excited, disappointed, or angry? If so, to what extent? Further, if you indicated that you are indeed in an emotional state, what is the cause of it? What specific event brought it about? Questions like these are being asked in hundreds of psychological studies conducted all over the world. They are also being asked in a similar form by economists and sociologists who assess life satisfaction, doctors who assess patient pain level or the presence of a psychiatric condition, marketing researchers who assess customer satisfaction, and many other professionals. Though the process by which people answer these questions may appear straightforward, it presumes that people actually know (a) that they feel an emotion, (b) what specific emotion they feel, and (c) why they feel it. In contrast, research in psychology reveals that, under some circumstances, people can be wrong about their emotions. Specifically, people can be in an emotional state without having any conscious awareness of being in that state. People can be wrong about the state they are in. And finally, they can be wrong and even unconscious about the causes of their emotional state. In short, the topic of unconscious emotion deals with the fundamental question of conscious access to and self-understanding of one's own emotional life. The remainder of this entry covers the historical background of this idea and gives several examples from modern research.

Historical Background

Historically, interest in the limits of emotional self-understanding and the relation of emotion to consciousness dates back to Sigmund Freud and several of his contemporaries. Freud speculated that people can sometimes be mistaken about what triggers their emotion. For example, a woman may believe that she is attracted to a man because of his professional achievements, but in fact she likes the man’s similarity to her father. Freud also speculated about the possibility of confusion about the nature of one’s own emotional state. For example, a man may believe that he feels angry at his partner for being late, whereas in fact he feels jealous—a feeling he either cannot identify, name, or perhaps admit. Finally, Freud wondered whether some emotions are sometimes “kept” from consciousnesses, such as son’s sexual feelings toward his mother or his homicidal anger toward his father.

Modern Research

Modern psychology has largely rejected Freud’s vision of psyche and his dramatic speculations. However, interest in those issues remains high, and empirical research, now armed with tools of psychology and neuroscience, shows that emotion, consciousness, and understanding can dissociate but usually in much less dramatic forms, as shown next. Note, however, that some of the following examples involve mild, undifferentiated emotional states (e.g., moods and changes in general positivity/negativity).
Mistaken Beliefs About Causes and Nature of One’s Emotion

Modern research shows that if a newly encountered person is similar to the participant’s significant other (e.g., sibling, parent, close friend), the participant will partially transfer (generalize) the traits of the significant other to the new person, without realizing that he or she is doing so. Several classic studies found that mood or arousal because of one source can transfer (spill over) to an irrelevant object. For example, participants aroused from riding an exercise bike rate pictures of members of the opposite sex as more attractive. Men who have just viewed an erotic picture rate a completely unrelated financial gamble as more profitable. In another example, participants called on a sunny day, and thus feeling happy, give higher ratings on a variety of judgments, including life satisfaction.

Emotional States Induced Without Awareness

It is possible to elicit mild emotional states by presenting stimuli completely without awareness. This is often done using subliminal presentations in which stimuli are presented in a way that prevents awareness. The method can involve flashing the stimuli very briefly (e.g., 5 milliseconds [ms]), presenting them outside the focus of attention (e.g., as flashes in the corner of the computer screen), masking (covering) the stimulus with another, more salient object, or typically, a combination of these methods. For example, participants who were subliminally flashed a large number of emotion-related words reported changes in a generalized mood state, without realizing what brought on that change. Further, repeated subliminal presentation of simple geometric figures (polygons) has been shown to lead to subtle mood enhancement, presumably because repeated stimuli elicit a warm feeling of familiarity. Another study showed that negative mood can be enhanced by subliminally presenting images of snakes and spiders to phobic individuals. Finally, many studies showed that very briefly presented emotional stimuli (e.g., faces) can lead to physiological and judgmental manifestations of emotion (e.g., as reflected in the activation of emotional brain systems and their bodily concomitants). Of course, with all these examples, it is important to remember that the stimuli used are very simple. This is important because there is a debate in psychology to what extent novel, complex stimuli can be processed without awareness and how strong and durable are reactions elicited by such stimuli.

Fully Unconscious Emotion

Perhaps the greatest controversy in psychology regarding unconscious emotion surrounds the possibility that people may sometimes not even realize that they are in an emotional state. There is some literature on dissociations between emotion and consciousness in hypnotic states. Some dramatic cases have been reported of people engaging in strong emotional acts while in a state characterized by lack of full consciousness (e.g., cases of “sleep murder,” “sleep intercourse”). Empirically, however, this kind of “fully unconscious emotion” has been a difficult topic to study because of the impossibility of completely excluding that a person did not feel something or wasn’t somehow aware. There are also limits on generalizing from cases that may involve psychopathology. Still, there are some intriguing results from a series of studies on typical college participants. Those studies unobtrusively exposed participants to several happy or angry subliminal emotional facial expressions. After that emotional induction, participants were asked to perform some emotion-related behavior (e.g., drinking a novel beverage, making a gamble). Participants were also asked to report their emotional state. The emotion state was also monitored using psychophysiology. Interestingly, in those studies the ratings of conscious feelings were unaffected by subliminal faces. Yet participants showed changes in their emotional behavior. For example, they consumed more of the beverage after happy rather than after angry faces. They gambled more after happy than after angry faces. Further, the psychophysiological measures indicated that the emotional faces elicited emotion-appropriate changes in the bodily state (e.g., more smiling and less startling to loud noises, etc.). As such, these results suggest that, at least in some circumstances, one can induce an emotional state that drives a person’s physiology and overt behavior without that state giving rise to conscious, reportable feelings. In short, there may indeed be fully unconscious emotion.

Piotr Winkielman

See also Consciousness and the Unconscious; Emotions and Consciousness; Subliminal Perception; Unconscious Perception
UNCONSCIOUS PERCEPTION

If you are reading this, you are probably experiencing conscious perception. Is it possible that you could understand this sentence without conscious awareness? If so, this would constitute unconscious perception. Consciousness per se has been difficult to capture operationally and scientifically, but progress has been made recently, at least in terms of neurological correlates of subjective awareness. Delineating the absence of consciousness has been fraught with methodological and theoretical challenges, but subjective unawareness on the part of the human perceiver has become scientifically legitimized. Namely, when there is any measurable change in one's experience, thoughts, or actions as a function of current external events juxtaposed with an absence of awareness of the events, then unconscious perception has occurred.

Different Types of Unconsciousness

Although Sigmund Freud is generally given credit for raising consciousness about unconsciousness, his brand—the psychoanalytic unconscious—is only one aspect. While Freud’s conception of the unconscious deals primarily with appetitive urges and motivations, more recent work by researchers including John Bargh, Ap Dijksterhuis, and John Kihlstrom has focused on—and found empirical evidence for—cognitive and social processes that operate at an unconscious level. Many of the cognitive and social processes involved in unconscious perception have been revealed to be surprisingly sophisticated and complex and are discussed below.

Terminology

A variety of terms have been used to describe perception without awareness: unconscious perception, nonconscious perception, subception, implicit perception, and subliminal perception. The term limen (a root of subliminal) implies a threshold for consciousness. Philip Merikle and Jim Cheesman (1986) have most effectively defined this boundary as a subjective threshold, or “the level of discriminative responding at which observers claim not to be able to detect perceptual information at better than a chance level of performance” (p. 42). However, since the subjective threshold is based entirely on a perceiver’s self-report, it should be combined with an additional criterion—performance that is qualitatively different during aware versus unaware episodes—which in turn permits distinguishing conscious from unconscious processes. Thus, the term implicit perception has recently gained popularity among cognitive neuroscientists, both because it avoids the logical problem created by the term subliminal (i.e., evidence suggests the existence of a continuum along the conscious-subconscious spectrum, not simply a threshold) and, according to John Kihlstrom and others, because subliminal perception is just one subcategory of unconscious perception. In particular, we shall see that unconscious perception covers a much wider range of phenomena than the narrower term subliminal. A PsycINFO search (June 2010) revealed 1,932 citations concerned with unconscious perception, in contrast to only 636 involving subliminal perception.

Unconscious Processes Versus Unconscious Stimuli

Information can be unavailable to consciousness for many reasons. On the event side, the stimulus can be uninterpretable or not sensed at all because it is too degraded (either optically or aurally distorted), too faint (too dim or too soft), or too brief (e.g., presented for 2 milliseconds). (The preceding are examples of an absolute threshold, a concept in psychophysical scaling.) Alternatively, the observer can be the cause of the unconscious processing. A person can be outright unconscious (e.g., asleep or under anesthesia), unaware because of concurrent attentional demands, or suffer from a neurological
condition that precludes conscious perception (e.g., blindsight, discussed below).

By definition, subliminal phenomena are restricted to stimuli that are either extremely degraded or presented so briefly that nothing meaningful can be perceived. For example, a word flashed on a computer screen for 1 millisecond (ms) is experienced as no more than a flash of light; although there is a sensation, neither word nor even letters are perceived. Nevertheless, even such stimuli can be shown to have been processed (“perceived”) subconsciously.

How can a stimulus be unconscious? It cannot, of course. But a person can be unaware of a stimulus or at least unaware of its identity. For example, in one typical experiment by John Seamon, Richard Marsh, and Nathan Brody (published in 1984), irregular 8-sided polygons were presented for 2, 8, 12, 24, or 28 ms. Polygons presented for 2 to 8 ms were indeed processed, as evidenced by subsequent “correct” affective judgments: Compared to nonpresented items, previously presented items were “liked better.” In contrast, conscious recognition (i.e., subjects had to determine which polygons had been presented previously) required presentations of at least 12 ms to achieve above-chance levels of performance. These findings demonstrate two principles: (a) Very briefly presented stimuli are (or at least can be) processed unconsciously, but (b) conscious access is available only for stimuli presented for somewhat longer intervals. (In this study, the threshold was presumably somewhere between 8–12 ms.)

Sensation, Perception, and Memory

Although cognitive psychology allocates these three processes to separate pigeonholes, of course they are all related. Indeed, the three phenomena normally function seamlessly and often subconsciously. Sensation can be defined as the raw input of external stimuli and involves processes by which sensory receptors and the nervous system receive and represent environmental stimulus energies. Although we are typically aware of sensory inputs, elementary sensation can occur without awareness. For example, the pupil of the eye can respond to light, and galvanic skin responses or event-related potentials in the brain can be recorded independent of concomitant awareness.

In contrast, perception—in its quintessential definition—typically requires some cognizance of meaning. Perception organizes and interprets sensory information, making it possible to understand the meaning of objects and events. A compelling distinction between sensation and perception (and between conscious and unconscious perception) can be illustrated by the case of a rare neurological syndrome, prosopagnosia (sometimes referred to as “face blindness”). A person with prosopagnosia can have perfect perception of individual facial features (e.g., eyes, nose, lips), but cannot synthesize those details appropriately to accurately recognize a well-known friend or spouse. It is unclear whether individuals with prosopagnosia have unconscious perception of familiar faces or whether such perception is simply neurally disconnected. (Once a familiar but visually unrecognized person speaks, that person can be identified immediately.) Foreign language provides another example: The spoken phrase tudo azul can be heard (sensed) by any normal listener as human speech but only perceived as a meaningful phrase by someone who speaks Portuguese.

Memory also affects perception, even (or perhaps especially) when such memory is unconscious (known as implicit memory). For example, even though subjects could not consciously remember pictures they had seen 17 years previously (some could not even remember having participated in the experiment!), David Mitchell found that implicit memory enabled perception of the corresponding 17-year-old picture fragments. In our everyday experiences, sensation, perception, and memory blend into a seamless process. However, the remainder of this entry focuses on events in the observer’s current perceptual field.

A Brief History of the Unconscious

Within experimental psychology, unconscious perception is one of the most venerable topics. Its primordial status can be traced to 1884, when C. S. Pierce and J. Jastrow published the first empirical report “on small differences in sensation” (without awareness). The concept of unconscious perception, however, predates empirical research. For example, in 1867 Hermann von Helmholtz theorized that that perception was dependent on “unconscious inference.”

Unconscious perception has been one of the most controversial topics in experimental psychology (a “checkered past,” in John Kihlstrom’s words).
Skepticism surrounding its existence (and accompanying methodology) has contributed to its controversial nature. Interest in unconscious phenomena has waxed and waned for over a century, both among scientists and the hoi polloi. The waning of empirical research on subliminal perception in particular may be traced to a famous but highly controversial movie theater “study” by James Vicary in 1957, in which he claimed that popcorn and Coca-Cola sales were increased by subliminal ads. Indeed, the 1960s saw only 8% of the studies in this area, but the number of studies picked up in the last three decades (71% of all subliminal studies). In contrast, the number of studies dealing with the more general concept of unconscious perception (including both cognitive and social processes) has shown a steady rise, with the count nearly doubling since 2000. Indeed, research in unconscious perception has mushroomed recently, and a 2005 volume by Ran Hassin, James Uleman, and John Bargh proclaimed the current era as “the new unconscious.”

Controversy

What made the topic of unconscious perception controversial? In spite of earlier empirical research documenting the phenomenon, the publicity surrounding Vicary’s hoax coupled with some failures to replicate, along with the behaviorists’ demonization of mentalism, made subliminal perception scientifically improper. Even following the “cognitive revolution” spurred by Ulric Neisser’s 1976 *Cognition and Reality*—where mental phenomena were restored to scientific legitimacy and intellectual respectability—the emphasis in cognitive psychology remained grounded in conscious experience, and the unconscious remained taboo. The zeitgeist subsequently shifted, due in part to John Kihlstrom’s 1987 “Cognitive Unconscious” article in *Science*, which restored the respectability of this field. Namely, the unconscious was no longer restricted to Freudian concepts or to pseudoscientific claims about subliminal advertising.

In addition to the controversial (pseudoscientific, cargo cult status) and methodological issues surrounding subliminal perception in particular, were fears foisted on the public by nonscientists claiming that advertising agencies were using subliminal techniques to effectively alter consumer behavior and endanger the American culture via sexual imagery (popular books by Wilson Bryan Key in the 1970s, 1980s, and 1990s, and by August Bullock in 2004). In spite of the negative publicity, however, legitimate research on unconscious phenomena in cognitive and social psychology finally picked up.

Empirical Evidence and Recent Research Topics

Hundreds of studies (nearly 1,300 since 1990 alone) have revealed—according to most researchers—unambiguous evidence for unconscious perception. As John Kihlstrom put it in 2008, the evidence “satisfies all but the most determined critics” (p. 587). Thanks to rigorous methodological developments, even the critics have come around, with only some particular phenomena excepted. For example, Anthony Greenwald and his colleagues demonstrated conclusively that “subliminal message self-help” audio media do nothing for consumers beyond standard expectancy (placebo) effects. For example, while products claiming to promote weight loss or to improve memory or self-esteem produced no such effects, audiotapes that were systematically mislabeled did affect consumers’ perceptions solely according to what was printed on the label. In a series of studies, John Vokey and Don Read investigated claims about “backmasking”—backward speech presumably embedded in rock music intended to promote evil behaviors—and found no evidence that subjects could process backward speech at any level. (Note that backmasking has no relation to *backward masking*, a rigorous experimental technique for eliminating conscious perception in visual tasks.)

In any case, the predominant current view is that unconscious perception is a viable, replicable, and scientifically respectable phenomenon. According to Sid Kouider and Stanislas Dehaene, the only controversy remaining has to do with the depth of processing of “invisible stimuli.” In other words, are unconsciously perceived stimuli processed only at a physical level (shapes, letters, or partial components of words) or at a complete and meaningful semantic level? The jury is not yet in on this question; evidence exists for both scenarios.

Stimuli That Are Not Consciously Perceived

Evidence that stimuli blocked from conscious awareness can be perceived comes from
process-dissociation procedures, neurological correlates, evaluative and affective effects, semantic priming effects, social judgments, perception under surgical anesthesia, behavior, and neuropsychological dissociations, just to name a few. Work by Anthony Marcel in the 1980s set the standard for rigorous laboratory demonstrations of unconscious perception in word recognition paradigms. Marcel used the technique of pattern masking, in which a mask presented immediately after a briefly presented word prevented observers from seeing the word. Nonetheless, subsequently presented words that were semantically related to the unperceived primes were facilitated. More recent work has employed “sandwich masking,” in which patterns are presented both before and after a brief stimulus, making the possibility of conscious perception extremely unlikely. The mere exposure effect championed by R. B. Zajonc includes affective preferences for stimuli experienced without awareness—even prenatally—even though the same stimuli cannot be consciously recognized.

Although most of the research on unconscious perception has involved vision, recent work has included unconscious perception in the realms of olfaction and audition. Unconscious auditory perception has been investigated not only in the laboratory, but also in patients under general anesthesia. In the latter studies, the stimuli are presented at an audible level (typically via headphones), but the perceiver is unconscious during input. After an early study by B. W. Levinson in 1965, this field was quiet, but with the advent of work on implicit memory, research picked up in the 1990s. In 2007, Jackie Andrade and Catherine Deeprose found at least two-dozen studies that included the Bispectral Index, a rigorous indicator of the “depth of anesthesia” (i.e., is the patient actually unconscious?). They concluded—with careful emphasis on studies that had controlled for “undetected awareness” on the part of patients undergoing surgery—that some degree of unconscious perception occurs during anesthesia. That perception (measured by implicit memory) can occur when patients are unconscious is so accepted that a recent article explored the possibility of “sex differences in memory formation during general anesthesia” (none were found).

Regarding neuropsychological conditions, blindsight (first reported by Lawrence Weiskrantz in 1986) involves a phenomenon wherein individuals with damage to the striate cortex in the brain cannot consciously see objects; however, they can point toward or grasp these objects accurately when encouraged to reach out. A similar dissociation between conscious perception of facial features without corresponding recognition of a familiar person occurs in prosopagnosia.

**Unconscious Processes**

In very recent research, Ap Dijksterhuis and his colleagues have discovered many advantages of unconscious processes. Complex decisions (e.g., choosing an apartment, playing chess, predicting soccer matches, clinical diagnoses) can be made faster and even more accurately unconsciously than consciously. Dijksterhuis argues that unconscious thought is faster because it is not constrained by the capacity limitations of conscious processes. Some investigators have concluded that the brain “does not care” whether representations are conscious or not—that is, the representation is just there. Such processes have also been found to influence unconscious behaviors such as walking speed. For example, John Bargh and his colleagues have found that people walked more slowly after being exposed to words related to “elderly” concepts implying slowness.

**Applications and Conclusions**

Regarding advertising, recent work has focused on “using and abusing” subliminal stimulation in this field, showing that under very specific circumstances, it is possible to influence consumer choices. For example, Dijksterhuis and his colleagues found that subjects can be stimulated to increase their subjective ratings of hunger and thirst, can be made to actually eat or drink more, and that even brand choices can be influenced. Some investigators argue that the evidence is strong enough to say that commercial applications of subliminal stimulation can work in principle and that such phenomena should not be treated as a myth unworthy of investigation. On the other hand, there has been abuse of popular belief in subliminal stimulation by marketing self-help audio media (“I feel fantastic,” “Let’s stay away from pizza,” etc.); research has repeatedly failed to find any benefit. However, some research suggests that visual subliminal perception may be more effective than auditory subliminal perception.
If true, a probable explanation is that information processing capacity is magnitudes greater for visual than the auditory modality. However, subliminal auditory stimuli can be processed—even across to the visual modality—but backward speech cannot be perceived at any level.

In conclusion, there are a number of good reasons for continuing to investigate unconscious perception, beyond the purely scientific interest of knowledge for its own sake. There may be potential beneficial use (improving health), and it may be important to know when there is potential for consumer abuse.

David B. Mitchell

Author’s Note: Research for and writing of this entry was facilitated by grants from the Foley Family Foundation and the WellStar Foundation.

See also Anesthesia and Awareness; Attention and Consciousness; Blindsight; Consciousness and the Unconscious; Implicit Memory; Neural Correlates of Consciousness; Perceptual Consciousness and Attention; Subliminal Perception

Further Readings
Visual imagery refers to the processes through which people create “mental pictures” that they can inspect with their “mind’s eye.” These mental representations resemble actual depictions both subjectively and functionally and play an important role in remembering and problem solving. More broadly, they can, by virtue of their content, powerfully guide the flow of thought.

The Subjective Qualities of Visual Images

Visual images can be (and usually are) created in the absence of an actual visual stimulus—and so one can create an image of (or “visualize”) an elephant even if none is nearby. One can also create images that alter things that are in view (and so one could, for example, imagine this page with all the words printed in green ink). More ambitiously, one can create images of things that do not exist at all (e.g., an image of a unicorn). In some cases, images are created deliberately (and so someone can, if they choose, call up an image of a beautiful sunset); in other cases, the images arise spontaneously (perhaps triggered by someone else’s mention of a sunset).

Visual images are not hallucinations—the person experiencing the image can tell that the image is “in their head,” and not a real sight. Nonetheless, there is a strong subjective resemblance between visual images and actual sights. This is reflected in the way people commonly talk about their visual images, and references to mental pictures or the mind’s eye have been common at least since Shakespeare’s time (e.g., Hamlet, Act 1, Scene ii). It is noteworthy that people feel these terms are apt; this is a strong indication that the conscious experience of having an image does resemble the experience of seeing. More specifically, the imaged object or scene seems to be “viewed” from some particular vantage point and is typically “seen” against some background; objects in the image have colors, shapes, and surface textures that are immediately “visible” and so on.

Experimental Studies of Visual Images

Here as elsewhere, though, researchers are cautious in how they interpret these self-reports on conscious experience. However, the self-reports can be corroborated via the appropriate experiments, and the data make it plain that visual images do share many functional properties with actual sights, as one would expect based on the self-report. For example, participants in one study were first asked to memorize a map of an island, including the location of several landmarks on the island. With this done, participants were asked to create a mental image of the map and then to “scan” their image from one landmark named by the experimenter to another. When these scanning times were carefully recorded, they showed a strong linear relationship between the time needed for each scan and the distance between the relevant pair of landmarks on the original map. This result confirms that the image accurately represents all the spatial relations on the map—and so points close together on the map are functionally close on the image; points further apart on the map are far
apart on the image. In this fashion, the image seems truly to depict the layout (and thus all the shapes and patterns) in the scene that is being represented.

Experiments also indicate that images function as though they have an identifiable “view point,” just as an actual visual scene would. This is evident, for example, in the fact that aspects of the image that depict larger objects or objects that are at the front of the imaged scene are more rapidly accessible. Likewise, participants need time to “zoom in” on an image to inspect small details or to “zoom out” to survey the larger scene, suggesting that again, just like actual scenes, images are inspected from a particular “viewing distance.”

Visual images also respect spatial layout in another regard—in the pattern of transformations in the image. In many studies, for example, participants have been shown two forms and asked if the forms are the same shape but viewed from different vantage points, or actually different shapes. In these studies, participants seem to imagine one of the forms “rotating” into alignment with the other and then make their decision only after this imagined rotation. The imagined rotation itself seems to take place at a constant velocity, and so the time needed for these judgments is a linear function of the angular distance between the initial orientations of the two forms being compared.

The Imagery Debate

Across the 1970s and 1980s, there was heated debate about the results just described. In part, the argument hinged on a methodological point—namely, the notion that these response-time findings might be attributable to the demand character of the experiments (i.e., cues within the experimental setting that signaled to the participants how they were “supposed to” behave). More broadly, there was debate about the meaning of the experimental results. Thinkers such as Zenon Pylyshyn argued that the data were the result of experimental participants seeking to “simulate” the relevant activities, more or less as a mime might seek to simulate some action; this simulation, Pylyshyn argued, was guided by participants’ tacit knowledge about the relevant events in the world (rotation, straight-line movement, etc.). In response, Stephen Kosslyn and others argued that mental imagery relied on a specialized representation “medium,” and that the experimental data were revealing the properties of this medium.

The narrow concern about experimental demand was easily dealt with by appropriate experimental controls, and indeed, the response-time patterns remain the same when these controls are in place. The larger issue, an explanation in terms of simulation versus one in terms of an imagery medium, required a different type of analysis, and although some scholars regard the issue as still open, most researchers believe that the data reviewed in the next section argue powerfully for the notion of an imagery medium.

Neuroscience Studies of Imagery

What brain mechanisms might lie behind the behavior data described so far? In the view of many researchers, the various parallels between visual imagery and actual vision invite the hypothesis that these two activities rely on similar brain circuits, and several lines of evidence confirm this claim. Neuroimaging results indicate an enormous overlap between the brain sites activated during visual imagery and sites activated during ordinary vision. Likewise, brain damage often has parallel effects on imagery and vision. Thus, patients who (because of stroke) have lost the ability to perceive color often seem to lose the ability to imagine scenes in color. Similarly, if as a result of occipital damage patients have a blind spot in a particular region of visual space or some restriction of the extent of visual space, they are likely to have a corresponding limit on their visual imagery.

Further confirmation comes from studies that use transcranial magnetic stimulation to produce temporary “lesions” in visual cortex. Not surprisingly, this procedure causes a disruption of vision, but crucially, it also causes a parallel disruption in visual imagery.

Differences Between Visual Imagery and Vision

Even acknowledging these important parallels, however, there are also differences between visual imagery and vision and between mental pictures and actual pictures. For example, some discoveries that are easily made from a picture (the reinterpretation of an ambiguous drawing) are enormously difficult if the participant is relying on a mental image.
of that picture. According to some authors, this is because the image—as a mental representation—is accompanied by a “perceptual reference frame” that organizes the depiction, specifying the figure/ground organization, how the form is parsed, where the form’s top is located, and so on. This reference frame guides how the image is interpreted and so can place obstacles on image reinterpretation (or image-based problem solving of any sort). Pictures do not on their own have this sort of reference frame (the frame must be created by the perceiver), and so pictures (“unorganized depictions”) are more readily reinterpreted than images (“organized depictions”).

A further distinction between images and pictures is suggested by cases in which brain damage has disrupted someone’s vision but spared their ability to perform imagery tasks; the reverse pattern (disrupted imagery but intact vision) has also been observed. In addition, studies have often documented normal or near-normal performance on various imagery tasks from individuals who have been blind since birth—individuals who are unlikely to be relying on a picturelike mode of representation. Findings like these have led several researchers to propose a difference between visual imagery and spatial imagery. The former type of imagery yields a representation that bears a closer resemblance to a picture (and so visual images, but not spatial images, depict an object’s color or surface texture), and the processes needed to create visual images rely heavily on brain sites ordinarily used for actual vision. Spatial imagery, in contrast, relies more heavily on brain sites ordinarily used for guiding movements through space (both overt bodily movement and covert movements of attention). Presumably, individuals blind since birth rely on spatial imagery, not visual, and likewise, spatial imagery is what allows patients with disrupted vision (because of brain damage) to perform normally on many imagery tasks.

The distinction between visual and spatial imagery is also valuable in explaining a different point—namely, the ways in which (neurologically intact, sighted) individuals differ from one another in their imagery abilities. Both self-report and behavioral testing indicate that this variation is considerable: Some people report rich, vivid, visual images; some report no visual images at all. Some people perform well on paper-and-pencil tests requiring them to imagine folding pieces of paper or spinning forms; others perform much less well. Recent studies suggest that these individual differences need to be assessed separately for visual and spatial imagery (especially since self-report measures of imagery are powerfully shaped by someone’s strengths in visual imagery, while the paper and pencil measures often reflect someone’s ability in spatial imagery).

The Role of Visual Imagery in Cognition

Finally, what is imagery’s role within the broader context of mental processing? The answer has many elements. For some purposes, imagery is essential for remembering. (Imagine someone trying to describe a previously viewed face; that person may have no choice but to call up an image of the face and attempt to describe the contents of that image.) For other purposes, imagery may not be essential for memory but is nonetheless enormously helpful. We know, for example, that easily imaged words are easier to remember, and that deliberate attempts to form images of the to-be-remembered material usually aid memory. (Indeed, the use of imagery is a frequent component of many deliberate mnemonic strategies.) Imagery also seems to play a role in autobiographical memory, and memories of past episodes often take the form of images of those previous episodes.

In addition, imagery can play a role in problem solving. This is plainly the case when the problem involves spatial arrangement. (Imagine trying to decide whether a sofa, viewed in a store, will fit well in your living room. Most people would try to solve this problem by visualizing the sofa in place to “see” how it looks.) However, imagery also plays a role in other sorts of problems, including a variety of mathematical word problems. Moreover, casting any problem’s elements in terms of a visual image can shape the sequence of thoughts that come to mind in thinking through the problem. For example, thinking about your pet cat in terms of an image will make the cat’s appearance prominent for you, and this can call to mind other animals with a similar appearance. If you had thought about the cat without an image, the appearance might have been much less prominent so that some other set of ideas would be likely to come to mind. In this way, the mere step of casting the problem in terms of an image can guide the selection of available ideas, and this may have important consequences for the flow of thought.

It should be mentioned, though, that there are contexts in which imagery can be an impediment
Visual Masking

Visual masking refers to the reduced visibility of one stimulus, called the target, because of the presence of another stimulus, called the mask. As the generality of this definition suggests, visual masking is not a unitary phenomenon. Instead, a broad range of masking effects exists depending on the types of target and mask stimuli as well as their spatial and temporal relationships. This entry provides a classification of different masking types and illustrates their use in the study of vision and cognition.

Typology of Masking

Typically, the target and the mask are briefly presented (e.g., 10 milliseconds [ms]) and three types of masking occur according to their temporal order. Let us denote by stimulus onset asynchrony (SOA) the time delay between the onset of the target and the onset of the mask. When the target is presented before the mask (by convention negative SOAs), we have forward masking, and when the opposite holds, we have backward masking. The case when the target and mask are presented with SOA = 0 is called simultaneous masking.

The most basic form of visual masking, called masking by light, occurs when the mask stimulus is a spatially extended uniform field of light. Depending on the type of the target, masking by light can be divided into two subtypes: masking of light by light (the target is, like the mask, a spatially uniform stimulus but with much smaller spatial size) and masking of pattern by light (the target is a patterned stimulus such as an alphanumeric character). The more interesting types of visual masking occur in visual pattern masking, when a patterned target stimulus is masked by a patterned mask stimulus. From a methodological point of view, one can distinguish between three types of pattern masking: When target and mask do not overlap spatially, forward pattern masking is called paracontrast, and backward pattern masking is called metacontrast. When target and mask overlap spatially, if the mask does not contain any structural characteristics of the target, one has pattern masking by noise. On the other hand, when the mask has structural similarities to the target, one has pattern masking by structure. Figure 1 provides examples of different types of visual pattern masking.

Uses of Masking

Historically, the first formal use of visual masking dates back to the second half of the 19th century. It was mainly used as a tool to investigate the temporal evolution of perceptual processes. For example, researchers wanted to experimentally address questions such as the time it takes for a stimulus to reach the observer’s awareness and the phenomenal duration of a stimulus. Today, visual masking continues to be a method of choice in investigating not only...
Visual Masking

Figure 1  Types of masking according to spatial and informational content of the stimuli. Left: paracontrast and metacontrast; middle: masking by noise; right: masking by structure.

Figure 2  Left: Type A forward and backward masking functions. Maximum suppression of target visibility occurs at SOA = 0, that is, when the target and the mask are presented simultaneously. Right: Type B forward and backward masking functions. Here maximum suppression of target visibility occurs at SOA values different than zero.

Figure 3  Nonmonotonic masking functions can also exhibit more than one SOA value where target visibility reaches a local minimum. On the left, forward and backward masking functions with two local minima each. These are called bimodal masking functions. When there are more than two local minima, the masking function is multimodal, or oscillatory, as shown on the right.
problems pertaining to the temporal aspects of visual processes but also those related to conscious and unconscious information processing. Furthermore, a large amount of research is devoted to reveal the mechanisms of visual masking, and several models and theories have been proposed in the literature. Research in visual masking is voluminous, and the monographs referenced below provide an in-depth review of relevant literature and findings. In the following, a few examples will be presented to illustrate different uses of visual masking.

As mentioned above, one critical parameter in masking is SOA. The function that plots a measure of target visibility (e.g., brightness, edge completeness, shape, identity) as a function of SOA is called the masking function. The shape of the masking function depends on stimulus parameters as well as on the task of the observer. Nevertheless, masking functions can be categorized under two generic types: monotonic (also called Type A) and nonmonotonic (also called Type B) functions, as illustrated in Figure 2. Furthermore, the nonmonotonic functions can be classified as unimodal, bimodal, or multimodal (oscillatory) types (Figure 3).

While a complete understanding of neural processes underlying these masking functions remains a fundamental research question, the morphologies of these functions provide valuable insights into the temporal aspects of visual processes. For example, in Type A masking functions, one can interpret the SOA interval where masking effects occur as the “broad temporal window,” during which the processing of the target and the mask occurs.

From a practical viewpoint, many researchers use masks as basic components of their experimental design to control the duration or effectiveness of the processes generated by their stimuli: Neural responses to a brief stimulus presented in isolation may persist hundreds of milliseconds and generate ceiling effects in performance. If this brief stimulus is followed by an appropriate mask, the mask-generated activity will interfere with the processing of the target, thereby reducing the effective duration and performance to a desired level. Let us note, however, that the effect of the mask is not a total stopping of the target activity; instead, the mask typically interferes in a selective way with the target. For example, the mask may render the contours of the target invisible, yet observers can report the location of this invisible target without any difficulty. The selectivity of the mask has been used to infer relative timing of different processes. For example, by comparing metacontrast masking functions obtained by contour and surface-brightness judgments, Bruno Breitmeyer and colleagues suggested that processes computing surface brightness are delayed with respect to those computing boundaries. Visual masking is also a powerful technique to investigate conscious and unconscious information processing. The mask can render the target completely invisible, thereby eliminating it from the visual awareness of the observer. However, as mentioned above, this does not mean that all target-related processes are extinguished; instead, the processing of several attributes of the target continues to take place at unconscious levels. By measuring the effects of the unconscious target stimulus on other stimuli, recent research has shown that a variety of stimulus characteristics, such as form and emotional content, are processed at unconscious levels.

Haluk Öğmen and Bruno G. Breitmeyer

See also Auditory masking; Consciousness and the Unconscious; Subliminal Perception; Unconscious Perception

Further Readings

VISUAL SEARCH

Visual search is the act of looking for an item. The target of one’s search can be defined as an object (such as a person), a feature (such as a color), or an event (such as the sudden appearance of something). Visual search has two roles in the study of the mind: first, as a phenomenon to be understood, and second, as a method for revealing how other aspects of
the mind work. The entry describes a typical visual-search experiment and how it is analyzed, reviews aspects of visual search that are of interest in themselves, and discusses the application of visual search to the study of cognition and the mind.

Visual search takes time because more visual information is hitting the retina at any given moment than one can process. Attention is a fundamental cognitive mechanism that enhances the processing of some information at the expense of momentarily irrelevant information. One way to attend to some items is through overt attention—that is, eye movements. However, one can attend to a subset of a visual scene with covert attention as well, most commonly referred to as “seeing out of the corner of the eye.” Researches on visual search and attention thus provide insights about each other.

**Method and Analysis**

Visual search is a common activity in daily life and in many cognitive activities: looking for a word on a page while reading, trying to find a friend in a crowded room, examining an X-ray for anatomical abnormalities in a hospital, or searching the ground for fossils and artifacts. The scientific study of visual search often examines much simpler tasks that afford greater experimental control. Of primary interest for many scientists interested in visual search is why some search tasks are fast and efficient and others are slow and inefficient.

Figure 1 displays an example of the sort of task a participant in a visual-search study would commonly see. Here the target is the letter T among non-target letter Fs. A typical experiment would require the participant to look for the target and to press one button if it is present and another if it is absent; half of the trials would have a target and half would not. The items would generally appear to the participant and remain visible until the response. The speed (or reaction time) and the accuracy of the search would be recorded for each trial.

The data are summarized commonly in a plot of reaction time as a function of the number of items displayed. Other variables, such as presence or absence of the target, are plotted as different functions or groupings of data points in the graph. Each group of data points can be summarized by a slope and intercept of the linear function that fits those points. The slope provides information about the rate of visual search, and the intercept summarizes the amount of time required for all other aspects of the search task (such as the motor preparation for the response).

Most analyses of the reaction time in a visual-search task assume that accuracy has not been traded off to achieve the speed found. (A speed-accuracy trade-off is common and occurs when speed is increased but accuracy is low and compromised for speed, or when speed is decreased and accuracy is high and improved due to the slower reaction time.) Most reaction time analyses in visual search require high levels of accuracy in all conditions so that a speed-accuracy trade-off does not confound the results. Of course, some scientists examine the effects on accuracy rather than on reaction time, perhaps with a manipulation of how long the display is shown to the participants. Yet others manipulate both speed and accuracy as an analytical method.

**The Phenomenon of Visual Search**

One of the most fundamental questions about visual search is what makes some tasks efficient and others inefficient. There is no clear dichotomy between efficient and inefficient search tasks; however, there is a great degree of variability in the relative efficiency of search tasks to be explained. As noted in the previous section, the slope is often of primary interest for understanding the relative efficiency of visual search with one set of variables versus another (see Figure 2 for typical, but invented, visual-search data). A slope of zero would occur when the reaction time does not increase as more items are added to the display; this would be considered an efficient search task,
such as finding the large, white $T$ in Figure 1. As the slope increases, then the task is considered inefficient because the number of items in the display slows the search task with each additional item; an example would be finding the black $T$ in Figure 1.

A number of variations of the task can be made to observe the effects on speed and accuracy. Common variables that are manipulated include the following: (a) the set size—that is, the number of items to search through including the target (Figure 1 displays a set size of 48), (b) the defining feature of the target (the letter $T$ in Figure 1), (c) the reported feature of the target (typically, whether it is present or absent in the display; participants could also be asked to make a judgment about it, such as what color it is or what the orientation of the target is), and (d) the degree of similarity between the target and the nontargets and amongst the nontargets (a $T$ might be more difficult to find amongst $F$s than $O$s, for example).

A common distinction in many search tasks is whether the defining feature of the target is a single feature (such as a single color, like red) or a conjunction of features (e.g., a combination of color and orientation, such as a red and vertical target among vertical nontargets that are another color, such as green, and red nontargets at the other orientation, horizontal). However, any search task can be made more or less efficient by manipulating the similarity of the target and nontarget items, independent of the number of features that define the target.

Several models have been developed to describe the mechanisms responsible for the efficiency of a visual-search task. Anne Treisman, in her feature integration theory, proposed that items are first processed simultaneously in a preattentive stage as basic features (such as color and orientation). If the target is defined as a single feature, then its presence can be detected at this stage. However, if a target is defined as a conjunction of features, then each item must be attended to have its features bound together and identified, with each item processed successively. Each item is attended in a random order until the target is found or until all items have been attended. Feature integration theory is still highly influential even though some details of this theory came into question after some experiments found that a target defined by a conjunction of features can, under some conditions, be found just as efficiently as a target defined by a single feature. Jeremy Wolfe later proposed a model that is similar to the feature integration theory, called guided search. With guided search, Wolfe accounts for these exceptions by not having items attended in a random order but rather

![Figure 2](image-url)  
**Figure 2** Typical (but invented) data for a visual-search experiment; reaction time (in milliseconds) is plotted as a function of set size.
Applications: Visual Search as a Window to Cognition

Visual search is a common method used to understand the mechanisms of attention. For visual search to be efficient, some information must be more easily prioritized than other information. Bottom-up and top-down mechanisms guide attention to items of interest. Top-down mechanisms guide attention based on the information that the observer is looking for, such as the color or orientation of the target. Bottom-up mechanisms guide attention based on the information in the scene and the manner in which it is processed in the early visual system of the brain. Unique items—rendered salient due to a contrast in color or orientation when compared to surrounding items—provide strong bottom-up guidance for attention and under some conditions even capture attention. The bottom-up salience of the items and the top-down, target-defining information that the participant uses combine to create an attentional priority map that indexes the locations of items that are likely to be the target. By manipulating the features of the target and nontarget items and the number of items, among other variables, visual search can reveal the mechanisms of attention and the prioritization of information for further processing, such as object identification and memory consolidation. Many studies have examined this both in terms of covert attention (where the eyes remain motionless, fixated at the center of the screen) and overt attention (where eye movements to different items are monitored and analyzed).

Visual search not only reveals how some information is prioritized but also how momentarily irrelevant information is inhibited. For example, in searching for a T in Figure 1, it would not be useful to check each nontarget F location more than once. Research in visual search provides converging evidence for a mechanism that suppresses attention to previously processed locations (inhibition of return). Visual search thus serves as a model task for how information is prioritized and processed by the mind.

Also of key interest in the study of information processing is that of serial versus parallel mechanisms. The search rate, as revealed by the slope derived from reaction time as a function of set size, provides some indication of whether items can be visually processed in serial (i.e., successively) or in parallel (i.e., simultaneously). Many mathematical models of parallel processing can mimic serial models, however, so visual-search rates provide only partial, converging evidence for this issue.

Although these fundamental issues of the mind can use visual search as a tool, so can many applied tasks. Visual search has been examined in reading, driving, and X-ray examination, just to name a few tasks that depend on fast, accurate visual search. As more is discovered about the basic mechanisms of visual search, the design and training for these real-world tasks can be improved to better suit how the mind processes visual information.

Michael J. Proulx

Further Readings


Visual Working Memory

To understand visual working memory, it is helpful to consider human memory more generally. Human memory is not a single mental faculty or cognitive system; rather, the cognitive processes that we
collectively call memory are composed of a number of independent and specialized cognitive systems that encode and store information in different formats. The various taxonomies of human memory agree on two main distinctions: First, memory may be explicit, or declarative, in the sense that personal knowledge and previous experiences are consciously recollected or recognized, or memory may be implicit, or nondeclarative, in the sense of being expressed indirectly in behavior without accompanying conscious recollections of previous learning episodes. Second, a distinction is drawn between short-term memory, called immediate memory by early memory researchers, which has been proposed as the seat of consciousness and active processing and which is able to store limited quantities of information for limited periods of time—in the range of seconds rather than minutes—and long-term memory, which stores unlimited amounts of information for unlimited periods of time. These distinctions are not completely orthogonal, since short-term memory is coupled to explicit memory.

The theory of working memory was proposed 35 years ago by Alan Baddeley and Graham Hitch and later developed in a very influential book by Alan Baddeley, *Working Memory,* published in 1986. According to this theory, short-term memory is not a simple mechanism for passive storage of information; rather, it is a coordinated set of mechanisms that combines incoming information from sensory systems with information retrieved from long-term memory and consists of a central executive operating with the assistance of domain-specific verbal, and visual support systems. Broadly speaking, the concept of visual working memory refers to the short-term memory system, which stores information that enters the brain through the eye and is maintained and manipulated by the support system referred to as the *visual-spatial sketchpad.* The definition may be too broad, however, because visual information, such as written words or pictures of naturalistic scenes, is easily recoded and stored in verbal-memory systems. A more restricted definition of visual working memory is *the maintenance and manipulation of information represented as visual codes.* Since performance with complex, meaningful visual stimuli may be supported by both visual and verbal working memory systems operating in concert, much of the research in visual working memory has aimed at isolating the visual component of memory. This entry considers the scientific evidence for visual memory codes and the organization and capacity limitation of visual working memory.

**The Evidence for Visual Memory Codes**

The idea of visual representations or memory codes has been challenged by researchers who favor a unitary verbal memory system. There is, however, mounting evidence that verbal and visual information is handled in separate cognitive and neural systems.

Perhaps the strongest evidence for separate visual and verbal memory representations comes from so-called dual-task experiments. In these experiments, subjects are required to carry out two tasks simultaneously. For example, they have to remember a previously presented visual pattern while counting backward or making a spatial judgment. The results of such experimental manipulations show that the performance on visual working memory tasks is not impaired by a parallel verbal task, but it is impaired by a parallel visual task. Thus, concurrent visual processing tasks compete for processing resources, whereas verbal and visual tasks are processed independently. A complementary interference pattern is found for verbal working memory. This pattern of interference strongly suggests the existence of separate, parallel, limited-capacity visual and verbal working memory systems.

Further evidence that the brain stores information in terms of visual representations comes from the study of brain-injured patients. In the neuropsychological syndrome of *visual agnosia* the patient sees, but the visual world is meaningless. There may be nothing wrong with the perceptual process, per se, and the person navigates safely in the geographic environment and may even copy diagrams and drawings. However, there is no visual recognition and no visual memory, although other forms of working memory (and long-term memory) may be completely intact.

Modern brain-imaging techniques, such as functional magnetic resonance imaging (fMRI) that allow fairly local and precise brain activity patterns to be mapped during the execution of cognitive tasks, would appear to offer a unique possibility of isolating the visual versus verbal components of working memory. But isolation of specialized memory systems is not a simple and straightforward procedure,
both because cognitive tasks activate many cognitive and brain systems concerned with perceptual, attentional, and memory processing, and because most neural networks in the brain probably perform many processing tasks in parallel. Even if the performance on a specific visual memory task is not assisted by verbal memory, the simple fact that we identify a recognizable pattern that can be verbally classified and later remembered implies activation of verbal memory processes. However, imaging studies have demonstrated distinct areas localized to the occipital and temporal lobes of the brain, specialized for processing particular classes of stimuli such as objects and human faces. These regions are activated in visual memory tasks with these classes of stimuli but not in verbal memory tasks, suggesting that at least certain kinds of information are represented as visual codes. There is also evidence that some visual memory tasks recruit visual-processing regions in the early stages of the processing hierarchy, which indicate a close relationship between visual perception and visual memory.

The Visual-Spatial Sketchpad

The theory of working memory includes two support systems for manipulation and maintenance of sensory information. The verbal support system, referred to as the phonological loop, is easily understood; it may be thought of as the process of repeating information in silent (subvocal) speech. The second support system, the visual-spatial sketchpad, is less intuitively understood. It is not a purely visual processing mechanism but thought to combine information from visual, tactile, and haptic sensory channels. To a first approximation the visual-spatial sketchpad may be thought of as largely equivalent to the phenomenon of visual imagery. Visual imagery is the ability to produce “inner” images of previously seen persons, objects, and geographical scenes and to produce images of scenes that are simply imagined. Visual mental imagery tasks require subjects to retrieve information from long-term memory or to maintain visual information that is recently viewed and to perform cognitive operations on or make judgments of this information. The purely visual nature of visual imagery is supported by experiments showing that cognitive operations on visual mental images are governed by the same laws as similar operations performed on online visual images, and that visual imagery recruits many of the brain regions involved in online perception, even including some of the earliest brain regions in the visual process. This process is quite similar to task of the visual-spatial sketchpad. However, whereas visual imagery is conceived as conscious process of image generation, maintenance in visual working memory is not conceived as a conscious process in the same sense. Differences between the visual imagery and the visual-spatial sketchpad are also suggested by the finding that visual imagery is disrupted by task-irrelevant visual noise whereas maintenance in visual working memory is not disrupted by visual noise. Thus, despite similarities, visual imagery and the visual-spatial sketchpad are not identical.

Fractionation of Visual Working Memory

Recent dual-task experiments indicate that visual working memory may be composed of additional subsystems. It is now claimed that there is one subsystem that is partly specialized for processing objects and object properties such as color, orientation, and shape and a second system that is specialized for processing the spatial characteristics of visual displays. For example, it has been shown that when subjects have to remember information about object color while simultaneously performing a mental task that requires shape recognition, there is substantial interference between the two tasks, whereas no interference is observed between a task requiring memory for spatial position and a second, shape recognition task.

Evidence for a possible third component of visual working memory comes from a parallel line of research, on so-called perceptual memory. In classical visual working memory tasks, participants are required to consciously remember the stimuli to be compared, and the difference between the patterns is easily detected in simultaneous displays. For example, in a typical experiment, one object in the display may change in color, orientation, or shape, and what is varied is the number of objects in the display and/or the number of object features that are changed. This approach is governed by what some cognitive psychologists call a storehouse metaphor—how much is retained in memory? Perceptual memory research, on the other hand, is governed by a correspondence metaphor, investigating the fidelity of memory. How well are the details of the original
display retained, and what distortions of color, texture, and orientation are observed? In these experiments, the memory of elementary attributes of visual images—size, orientation, movement—is measured using psychophysical discrimination tasks where participants are requested to decide, or guess, if a test pattern has a higher or lower value than a previously shown reference pattern on the stimulus dimension in question, and a discrimination threshold is determined, representing the value of 75% correct guessing. The results from these experiments show, somewhat surprisingly, that information about such elementary attributes is not degraded or distorted but maintained with the fidelity of the sensory image during intervals up to two or three minutes. The storage is impaired by a concurrent processing task along the same dimension but not by a concurrent task along another dimension, and there is evidence that at least three stimulus dimensions may be stored in parallel. Thus, the pattern of interference mimics the interference pattern observed in the more traditional visual working memory tasks. However, the performance on this task does not require conscious recollection or recognition; it turns up in forced choice tasks and may be more implicit in nature. It is an open question whether this low-level memory system is part of the visual working memory system or a parallel memory system associated with implicit memory.

Capacity Limitation of Visual Working Memory

The capacity of visual working memory depends upon several factors, such as the type of information tested, the test procedure employed, the time allowed for memory encoding, and the discriminability of the items in the visual display. There are individual differences in working memory capacity, and there are age effects. Visual working memory performance in the adult population starts to decline around 60 years of age and decreases steadily toward older age, a trend observed for most memory tasks involving explicit but not implicit memory.

Until quite recently, most research on visual working memory has been concerned with the spatial aspect of visual displays, and comparatively little was known about memory for the properties of visual objects and visual scenes, such as their color, texture, and orientation.

Experiments investigating the object property aspect of visual working memory suggest that visual-object working memory is a unitary mechanism, with a capacity limitation that is fairly independent of object properties and complexity. To a first approximation, visual working memory may be conceived of as storing integrated objects, thus when remembering objects and their properties, it does not matter if they have different colors and orientations or if they share color and orientation. The capacity limitation is four objects; when the number of objects exceeds this limit, memory suffers. This does not mean that the visual working memory capacity is completely independent of complexity. While the magical number four pops up in most experiments, it is clear that several stimulus and encoding parameters determine the capacity of the visual working memory. Thus, if object complexity is high and encoding time is low, the estimated visual working memory capacity decreases, but if object complexity is low and there is unlimited encoding time, capacity increases but never exceeds the magical number four.

In summary, visual working memory consists of a set of limited-capacity, short-term memory mechanisms handling information transmitted through the eyes, operating in concert with other working memory systems. Because visually transmitted information may be coded in both verbal and visual representations, it is likely that few everyday visual short-term memory tasks will be solely reliant on visual working memory.

Svein Magnussen

See also Divided Attention and Memory; Memory, Neural Basis; Visual Imagery; Visuospatial Reasoning; Working Memory; Working Memory, Evolution of

Further Readings

Visuospatial Reasoning

Visuospatial reasoning is essential for survival. If we couldn't find our way home or get food into our mouths, life would be nearly impossible. Visuospatial reasoning occurs whenever we go beyond the information given to make inferences about what we see or where we are, as when we recognize a friend from a glimpse of her face or see a face in the moon or recognize a neighborhood from the architecture. Visuospatial reasoning is involved in more than recognizing things and places, it also is involved in imagining things and places and mental transformations on things and places. Mundane activities, such as catching a fly ball and packing a suitcase, as well as sophisticated ones, such as refining a model of plate tectonics or designing a museum, rely on visuospatial reasoning. More surprising, visuospatial reasoning underlies abstract thought, such as deciding whether a giraffe is more intelligent than a tiger or whether a conclusion follows logically from its premises. Visuospatial reasoning allows inferences about motion in space as well as inferences about things in space. Even small children can distinguish motion paths caused by actions of animate agents from motion paths caused by inanimate agents. Because visuospatial reasoning is mental and therefore not directly observable, research began only after the behaviorist grip on scientific psychology loosened and the cognitive revolution of the 1960s began. It began with work on images, visual mental representations, and mental transformations and then broadened to explore spatial mental transformations and created spaces. This entry follows that history.

Visual Images

Demonstrating the psychological reality of images entailed showing that they preserved visual properties of the objects they represented, such as size, shape, distance, orientation, and color. In that spirit, some studies have shown that the time to mentally detect or compare those properties in images reflects actual differences in size, distance, or orientation. Other studies have shown that judgments of imagined shape or color yield the same similarity spaces as judgments of perceived shape or color. This approach views imagery as internalized perception. That is, having an image is like having a percept, and performing mental transformations on an image is like perceiving changes in the world. Impressive support for the psychological reality of visual images and the transformations performed on them has come from brain studies that have shown that visual imagery relies on some of the same areas of the brain involved in visual perception.

Mental Transformations

It is common in analyses of cognition to distinguish mental representations of aspects of the world from mental transformations of those representations, a distinction analogous to data and operations performed on data. One of the more remarkable visuospatial mental transformations demonstrated is mental rotation: the time to judge whether a pair of figures at different orientations, such as a configuration of boxes or an asymmetric letter of the alphabet, is the same or whether mirror images increase linearly with the degree of tilt, as if the mind were rotating the figures into congruence at a steady rate. Mental rotation has a real-life role in identifying and imagining objects that are not upright. Mental size and shape transformations also seem to be in the service of object recognition and, like mental rotation, can be appropriated for the kind of creative thought that serves rearranging furniture, understanding how molecules fit together, or designing a building. These are mental transformations of objects in space. Another major class of visuospatial mental transformation is imagining reorientations of one's own body in space, a set of skills that plays roles in tasks such as navigating a crowd or a new neighborhood, dancing, and playing Frisbee.

The two major classes of spatial mental transformations, mental transformations of things in the world and mental transformations of our bodies in the world, are rooted in everyday perceptual experience of the world, consistent with the imagery-as-internalized-perception approach. Interestingly, both


classes of mental transformations appear to involve motor as well as spatial reasoning. When participants are engaged in mental rotation while turning a wheel, the fluency of mental rotation is enhanced by turning the wheel in the direction of rotation and hindered by turning the wheel in the opposite direction. Another example comes from recognition of bodies in motion. The stimuli for these studies come from films of participants dressed in black with lights attached to their joints, moving in various ways. Although all that is visible in the films is the movements of the lights, these movements are readily perceived by human observers as walking, dancing, or skipping. People can readily distinguish men from women and friends from strangers. Surprisingly, recognition of bodies in motion is better for our own bodies than for those of friends, even though we have little visual experience of our own bodies in motion. The fact that people recognize themselves better than they recognize friends suggests close interactions between motoric and perceptual processes, interactions thought to underlie speech understanding, motor learning, empathy, and other processes as well. Observing human movement is thought to resonate in the observer’s motor system; that is, the observer may mentally simulate the movements while watching them. Observers seem to recognize when the simulated pattern is similar to the pattern evoked by their own movement. Conversely, motor movements affect visual recognition; for example, recognition of graspable objects like cups is facilitated when the hand is held in the appropriate grasping movement.

**Spatial Mental Representations**

Visual images are one kind of mental representation, but there are other kinds of visuospatial mental representations. Imagery for action has also been discussed and can be used for mental practice of playing the piano or gymnastics. There is also imagery reflecting other senses, such as sound, touch, smell, and taste as well as imagery not based in a particular sense, such as imagery for emotion or space. Spatial mental representations are typically more abstract than a specific sensory system. Congenitally blind people can form rich and full spatial mental representations from other sensory information, especially kinesthetic and auditory. These senses also play a role in spatial mental representations of sighted people. Spatial mental representations of the world reflect both perceptions of the world and actions in the world. Thus, they differ for the different kinds of spaces that people inhabit. The spaces important for human behavior and interactions are the space of the body, the space around the body, and the space of navigation.

As mental representations, images have been characterized by fidelity to visual attributes of the world. In contrast, mental spatial representations reflect behavioral and hence functional aspects of the world as well as visual aspects. The visual features do not always align with the spatial and behavioral aspects, in ways that will be revealed below.

**Space of the Body**

In order to accomplish daily or extraordinary tasks, the body moves in the world. For the body, we have both insider and outsider—visual—knowledge. Proproceptive feedback from our own movements and the responses of the environment is crucial to coordinated movements but also seems to play a role in recognizing movement as well as movers, as was seen in the studies using point-light movers. In other research, people were shown pictures of bodies with parts highlighted and asked to verify whether the highlighted part is the same as a named part. The more functional, important parts—for example, head and hand, are relatively small. However, people verify the small, functionally important body parts faster than the large ones, in sharp contrast to research on imagery.

**Space Around the Body**

The space around the body is the three-dimensional space surrounding the body in reach of hand or eye. As people—and rats—go about the world, they keep track of the changing positions of the objects around them relative to their own bodies, forming and updating egocentric mental representations. Perceptual and behavioral aspects of the body and the world affect the organization of those representations. The body has three axes: an asymmetric axis from head to feet, an asymmetric front-back axis, and a more-or-less symmetric left-right axis. The encompassing world also has three axes, only one of which is asymmetric, the up-down axis of gravity. These aspects of the body and the world have implications for perception and behavior. Both
perception and action are primarily directed forward as the front/back axis separates the world that can be seen and manipulated from the world that is difficult to perceive and manipulate. Gravity affects where we are, hence what we see, and how we can move in and manipulate the world. It also affects how objects in the world look, change, and can be interacted with.

These properties of the body and the world shape the mental representations of the space around them that people form and update. In one research program, people read descriptions of “you” in an opera house or museum. The descriptions placed objects around “you”—above, below, front, back, left, and right—after which a computer repeatedly reoriented “you” to face different objects and queried the participant for the current positions of all the objects. In particular, when the character (“you”) was upright, objects located above and below the body—that is, at head and feet—were easiest to retrieve because of the double asymmetries of the body and the world. Objects in front and back were next easiest because of the strong body asymmetry, and objects to right and left were most difficult, because those body asymmetries are weakest. Variations of the experiment established the spatial mental representations from models, diagrams, or real life, yet the consequent mental representations were essentially the same.

**Visuospatial Reasoning**

**Space of Navigation**

People (and rats) not only build mental representations of where things are relative to their bodies, they also build representations of where the objects are relative to each other and the world at large, *allocentric mental representations*. Allocentric representations are needed for navigating the larger world from many different directions. These, too, are biased, a consequence of the way spatial scenes are encoded and represented.

Mental representations of large spaces, or *cognitive maps*, are an amazing feat of the mind. Three-dimensional spaces that are too large to be seen from any particular viewpoint can be assembled from navigation with a view from within, from descriptions, and from external maps and then mentally shrunk and turned into a two-dimensional representation with a view from above. Assembling the pieces depends on coordinating common objects and reference frames, a process that distorts judgment and memory so that geographic bodies are remembered as more aligned with each other and with their reference frames than they actually are. Consequently, people erroneously think that Rome is south of Philadelphia and that Berkeley, in the east bay, is east of Stanford, which is west of the bay but east of Berkeley. Actual navigation is situated in the world, which corrects at least some of these misconceptions.

**Created Spaces**

**Describing Space**

Great literature, such as novels and poetry, and small literature, such as tourist guides and instruction manuals, all rely on the power of language to create mental spaces. The studies of the space around the body showed exactly that: From descriptions, people form spatial mental models that can be mentally transformed, allowing inferences and information retrieval. From verbal directions, people can imagine a battle scene, arrive at their desired destinations, or put together a piece of furniture. Spatial descriptions require a perspective. In describing space, people spontaneously adopt either a route (egocentric or intrinsic) perspective, where landmarks are described relative to a traveler from the traveler’s right, left, front, or back, or a survey perspective, where landmarks are described relative to each other using an allocentric reference frame such as north, south, east, west. Just as frequently, people mix perspectives, often mi-sentence and usually without warning. When people learn limited environments from either of these perspectives, they form mental representations that are perspective free—that is, that allow answering questions from either perspective equally quickly and accurately.

**Space of Graphics**

People have long used their extraordinary visuospatial capacities to create cognitive tools, such as diagrams, pictures, maps, and graphs that increase the power of the mind by augment memory, information processing, and communication. These graphics reflect the spatial metaphors that support abstract thought. For example, proximity in space represents proximity on abstract dimensions, as in graphs and in expressions such as, “We’ve grown
apart.” Direction in space also has meaning, especially up and down, going against or with gravity. Going up, against gravity, reflects strength, goodness, and other positive qualities, which are typically plotted upward. These biases are evident in gesture and speech as well in reasoning: “Give me a high 5,” “She’s at the top of the heap,” and “He’s fallen into a depression.” And with that, we end: “Onward and upward!”

**Barbara Tversky**

See also Common Coding; Mirror Neurons; Spatial Cognition, Development of; Thinking; Visual Imagery; Visual Working Memory; Working Memory, Evolution of

**Further Readings**


**Voluntary Action, Illusion of**

What distinguishes a voluntary action from an involuntary one? That the former may be expressed at will does little to answer the question, because “wills” are difficult to observe in others or other species. For instance, whether a monkey looks rightward voluntarily or reflexively is indistinguishable to an observer. (The former may occur when a monkey is trained to look away from an attention-grabbing flash in the anti-saccade task; the latter may occur during the visual grasp reflex.) Thus, drawing a principled distinction between actions that are voluntary and involuntary is less than straightforward. This entry explains from a cognitive science perspective how voluntary action is intimately related to “suppressibility,” the skeletal muscle system, ideomotor processing, and the sense of authorship.

**Suppressibility**

It has been proposed by Richard Passingham that unlike involuntary actions (e.g., reflexes, automatisms during seizures, and unconscious actions during pathological states), voluntary actions are special in that they can be suppressed. Though involuntary actions may be counteracted, as when the patellar (knee-jerk) reflex is counteracted by contracting the leg muscles, such indirect control is different from the kind of direct suppression that occurs during voluntary action, as when one refrains from making a comment or dropping a hot dish. Hence, suppressibility serves as a useful behavioral index of voluntary action in humans and other species.

**Sense of Agency and Ideomotor Processing**

With respect to one’s subjective experience, however, voluntary and involuntary action could not feel more different from each other. As researched extensively by Daniel Wegner and colleagues, the subjective sense of agency associated with voluntary action is based on several high-level cognitive processes, including the perception of a lawful correspondence between action intentions and action outcomes. If one intends to flex one’s index finger and then it happens to flex, one is likely to believe that the movement was voluntary. One is unconscious of the sophisticated motor programs giving rise to the action outcome but is often aware of the perceptual consequences (e.g., perceiving flexing a finger). William James and contemporary “ideomotor” theorists such as Wolfgang Prinz propose that, for voluntary action, the conscious mind then uses such perceptual-like representations to later influence the generation of motor efference (the signals to the muscle fibers), which itself is an unconscious process. From this standpoint, when one deliberately performs an action, one aims to reproduce the same sensations that occurred when a similar action was done in the past. To effect a desired efference to the muscles, all the conscious mind has to do is pay a certain kind of attention to such perceptual-like representations. Then, following some sort of “go” signal and if not curbed or controlled by James’s “acts of express fiat” (i.e., exercise of veto), unconscious motor programs take care of the rest by meticulously activating the right muscle fibers at...
the right time, thus giving rise to the action. John Bargh’s classic research demonstrates that incidental stimuli, such as ambient objects in one’s environment (e.g., a dartboard) can unconsciously activate processes that then influence one’s behavioral dispositions (e.g., making one more competitive).

Matching intentions to outcomes also underlies the sense of agency in the mental realm. If one intends to imagine the Mona Lisa and then happens to experience the relevant imagery, one is likely to believe that the imagery arose voluntarily, even when the percept may have been caused by an experimental trick, as in the Perky effect. (In the Perky effect, experimental subjects are fooled into believing that they are imagining an image that is actually presented physically on a screen.) In such a way, experiments on “authorship processing” by Wegner and colleagues have demonstrated that subjects can be fooled into believing that they caused actions that were in fact caused by someone else. Together, the findings indicate that we experience agency when our intentions satisfy the causal principles of consistency, priority, and exclusivity: Our intentions should be consistent with, and be experienced at an appropriate interval prior to, the relevant action, and there should be no other available explanation for the action. It has been shown that experimentally manipulating these three factors leads to systematic distortions in the sense of agency.

Intentional Binding

In intentional binding, investigated by Patrick Haggard and colleagues, the perceived elapsed time between voluntary actions (pressing a button) and their consequences (hearing a tone) is shorter than the actual time elapsed. The action is perceived as occurring later than it did and the outcome as occurring earlier than it did, as if the two events were temporally attracted to each other. In this way, one also binds the actions and outcomes performed by others.

Beyond findings showing the illusory and malleable nature of the sense of agency, there are good arguments a priori that there should be no such thing as an undetermined free will, homunculus, or ghost in the machine: The premise in science is that there cannot be a thing (e.g., a will) that is undetermined by past events. The classic research by Benjamin Libet corroborates this deterministic view. Libet instructed subjects to move their hand at will and to indicate when they experienced the conscious urge to perform the action. Although the consciously reported urge to move came about 200 milliseconds (ms) prior to the movement, detectable neural events began approximately 550 ms before the movement and, importantly, 350 ms before the reported urge to move. In short, unconscious neural activity preceded the onset of the conscious urge in a predictable manner. More recently, the research group led by John-Dylan Haynes discovered that the outcome of a willed decision can be predicted by unconscious brain activity occurring up to 10 seconds before the decision is made.

Skeletal (“Voluntary”) Muscle

Another feature of voluntary action is that it is limited to skeletal muscle effectors. Other effectors (e.g., smooth muscle) cannot be influenced by it, at least not directly. It has been proposed that skeletal muscles are “voluntary muscles” because they are directed by actional systems in the brain that, in order to influence skeletomotor action collectively, require conscious states to crosstalk with each other. Thus, consciousness is intimately related to voluntary action. Unlike involuntary actions (e.g., reflexive inhaling, the pain-withdrawal reflex), voluntary actions can be construed as a form of integrated action, which occurs when multiple action plans that could normally influence behavior on their own (when thus activated) are coactivated and trying to influence the same skeletomotor effector, such as when one holds one’s breath, refrains from dropping a hot dish, or makes oneself breathe faster. As Richard Passingham noted, suppression is an archetypal integrated action.

In conclusion, the experience (or illusion) of voluntary action can be indexed by suppressibility and characterized by its temporal properties (e.g., intentional binding), its link to the skeletal muscle effector system, the nature of the representations directing it (the conscious perceptual-like representations described in ideomotor theories), and the causal principles that furnish it with a sense of authorship.

Ezequiel Morsella, Margaret T. Lynn, and Travis A. Riddle

See also Access Consciousness; Action Slips; Automaticity; Common Coding; Freedom of Action; Phenomenology of Action; Philosophy of Action; Preconscious Free Will
Further Readings


**Williams Syndrome**

Williams syndrome (WS) is a neurodevelopmental disorder caused by the deletion of some 28 genes on one copy of Chromosome 7. The deletion gives rise to distinctive facial features and affects the developing body, brain, mind, and behavior. WS occurs in roughly 1 in 15,000 live births. Compared to more common disorders, such as Down syndrome, why would such a rare syndrome be of interest not only to psychologists but also to philosophers, neuroscientists, linguists, computer scientists, molecular biologists, and educationists? The reason lies in the promise that WS seemed to hold for substantiating a popular view of the mind/brain, namely, that it is composed of innately specified, independently functioning modules. Indeed, initial reports suggested that, despite low IQs and seriously impaired visuospatial and numerical cognition, individuals with WS had intact language and face processing. Was this not proof that language could develop independently of general intelligence? After more in-depth analyses of the phenotype, it became clear that WS was far more complex than researchers originally thought. In reality, far from illustrating the juxtaposition of intact and impaired modules, WS turned out to be a model of the extreme complexity of genotype/phenotype relations and of how domain-general deficits in infancy could cascade over developmental time to result in seemingly domain-specific outcomes in adulthood. In the following, some of the latest genetic, brain, and cognitive findings on this intriguing syndrome are presented, showing how WS continues to offer insights to all those fascinated by the human mind.

**Genotype/Phenotype Relations**

Early molecular research identified partial deletion patients (PD) with only two genes missing in the WS critical region: elastin (ELN) and LIM-kinase1 (LIMK1). ELN is important for elasticity of skin, lungs, and blood vessels and seemed to explain the WS arterial problems and facial dysmorphology. The PD patients also presented with visuospatial deficits implicating LIMK1, which is expressed in the brain, in these cognitive impairments. However, studies of other PD patients revealed neither facial dysmorphology nor spatial impairments despite deletions of ELN and LIMK1. Current research on PD patients with larger deletions, as well as animal models, indicates that four genes at the end of the deletion are those critical for the full WS cognitive and physical phenotype. Interestingly, these are all “transcription factors”; that is, they regulate numerous other downstream genes, suggesting that any one-to-one mapping between specific genes and specific cognitive outcomes is highly unlikely.

**The Williams Syndrome Brain**

The WS brain is clearly not a normal brain with parts intact and parts impaired. Its volume is only 80% of typically developing brains, with widespread differences across brain regions: particularly small cerebrum but average cerebellum; abnormal size and shape of the corpus callosum, the central sulcus, and the orbitofrontal cortex. Parietal regions have
reduced gray matter with abnormal neuronal layering, orientation, density, and size. Where normal brains become increasingly lateralized (hemispherically specialized) over time, the WS adult brain continues to process stimuli bilaterally, also revealing abnormal connectivity between the orbitofrontal cortex and the amygdala. Note that our knowledge of the WS brain results from research on adults and not on the developing WS brain, which would be informative regarding how differences compound or are compensated for by other networks over developmental time.

The Williams Syndrome Cognitive Profile

Individuals with WS present with an average full IQ of 56, but usually verbal IQ is higher than performance IQ. It is this uneven cognitive profile that has attracted attention. However, even within the domain of language, relative strengths in vocabulary can coexist with serious impairments in pragmatics and complex grammar. Visuospatial abilities also display peaks and troughs, with particular deficits in constructional and strengths in perceptual abilities.

One domain of particular interest is WS face processing, because on standardized tasks scores fall within the normal range. Could this be an example of a preserved function? In fact, brain imaging and cognitive studies have shown that the apparently normal performance stems from different brain processes. Again, a developmental approach is critical. For example, when studying spatial frequency biases in face recognition over development, children with WS demonstrate an adultlike bias much earlier than typically developing children, pointing to a less flexibly developing system. WS adults use more featural than configural strategies when processing faces. WS illustrates how, even when scores fall in the normal range, it is vital to probe the underlying brain and cognitive processes.

Conclusion

What, then, can be learned from the study of Williams syndrome? Not that WS will be a model of direct-gene behavior or cognition-brain region mappings. Rather, WS illustrates the complexities of the human mind/brain. While significant advances have been made in the genetic, brain, and cognitive domains, it will be critical to bring these complementary levels of analysis together. It is also essential to understand how having a syndrome like WS alters the environment in which a child develops and thus to study disorders within the full context of their developmental trajectories. Cross-syndrome studies will help identify which traits are truly syndrome specific. Williams syndrome provides an ideal model for the developmental study of how gene expression, brain, cognition, and environment are integrated to give rise to both the typical and atypical human mind.

Hayley C. Leonard and Annette Karmiloff-Smith

See also Autism; Face Perception; Language Development; Spatial Cognition, Development of

Further Readings


Wisdom of Crowds Effect

Psychologists have historically conceived of crowds as suppressing individuality. Recently, an alternative vision of crowds has emerged: Each person potentially brings unique insights, which if combined properly can make the crowd a better decision maker than most individuals. This entry will
discuss the conditions under which crowds are wise, whether individuals acting alone can mimic the effects of a crowd, as well as psychological biases that may prevent people from taking full advantage of what crowds have to offer.

Published demonstrations of the wisdom of crowds effect go back to the early 20th century. In one early study from the 1920s, students estimated the temperature in a classroom. When the estimates were averaged together, the resulting group answer was more accurate than the estimate of a typical member. Although early authors attributed the result to some mysterious group property, the statistical underpinning of the effect is now generally understood: A large sample of imperfect estimates tends to cancel out extreme errors and converge on the truth. Subsequent research demonstrated that simple algorithms that weight people equally, such as averaging, often compare favorably to more sophisticated statistical methods of combination. The literature on aggregation was reviewed by Robert Clemen in a 1989 paper in the *International Journal of Forecasting* and more recently by J. Scott Armstrong in his 2001 book *Principles of Forecasting*. The power and simplicity of averaging was also featured in James Surowiecki's 2004 best-selling book *The Wisdom of Crowds*. The logic of tapping diverse perspectives extends to many tasks, including identifying decision objectives, generating alternatives, and choosing among alternatives.

### Conditions for Crowd Wisdom

To take full advantage of collective wisdom, groups should be composed of people with topic-relevant knowledge or expertise. As important, the group needs to hold diverse perspectives and bring different knowledge to bear on a topic. Diversity helps because any given perspective is likely to be wrong. People who share a perspective will all be wrong in the same way (e.g., numerical estimates that all over- or underestimate the truth), in which case there is little benefit gained from a crowd. For numerical estimates, the benefit comes when errors “bracket” the truth and cancel out. Interestingly, diversity is so valuable that one can still benefit from averaging when individuals differ greatly in accuracy. In short, knowledge and diversity are the reasons that crowds are often wise.

Differences in perspective (and bracketing) are created both through who is included in the group—when people have different experiences, training, and judgment models—and through process—when ideas are formed and expressed independently from the ideas of others. The importance of process is illustrated by a result in the brainstorming literature. In the classic approach to brainstorming, people generate ideas face-to-face, and build on one another’s ideas. However, these interacting groups perform less well—in quality and quantity of alternatives—than noninteracting groups. Although exposure to others’ perspectives benefits individuals, over time it can lead people to think more alike, and diversity of perspective is lost.

### Can a Person Be a Crowd?

An intriguing recent area of research has extended the logic of the wisdom of crowds to individuals. It turns out that people can achieve some of the benefit of a crowd by digging deeper into their own minds. The key insight is that people typically rely on only a sample of the evidence available to them at any given time. But what if people had a reset button, so that they could retrieve facts from memory anew or handle the same facts in a new way? Simply asking people to answer again does not work; people will inevitably anchor on their initial opinions. There are at least two effective ways to break this anchoring effect, both illustrated in recent papers in *Psychological Science*. First, Edward Vul and Hal Pashler showed that people can be freed from their original answer by delaying a second answer. With the time delay people may forget their initial perspectives and think about the problem differently. The second approach, developed by Stefan Herzog and Ralph Hertwig, is to ask people to assume that their first answer was wrong and to answer the question again. Overall, averaging two opinions from the same person using either time delay or “assume you’re wrong and answer again” improves performance by about half as much as averaging across two people.

### Psychological Obstacles to Crowd Wisdom

Given that crowds are often wise (including the crowd in the mind), an important question for psychology is whether people make the best use of knowledge that is distributed across perspectives. In general, one can conceive of people using advice from others in three stages: People first collect
opinions, then combine the opinions into a judgment or belief, and finally hold this belief with a certain degree of confidence. When it comes to making the most of diversity, people fall short at all three stages. First, people do not uniformly seek out additional opinions. When they do, they often do not seek diversity. Instead, they collect opinions from relatively homogenous sources that share a common perspective, either because they seek confirmation or because similar others are more proximate. For example, a doctor may talk to a colleague with the same specialty or training, and an economist may discuss a forecast with someone who shares the same theoretical assumptions. Second, people combine fewer opinions than they should. One reason for this is that many people have incorrect intuitions about averaging, believing that it locks in the accuracy of the average judge in a crowd. Another reason is that people are overconfident in their ability to identify expertise and consequently “chase the expert” by selecting the single opinion they believe to be most accurate. Even with a larger group, people may focus on themselves or on just a few judges and miss out on the wisdom of the rest. In a 2009 article in *Management Science*, Albert Mannes showed that neglecting others comes at a high price in large crowds. Third, as shown by David Budescu and his colleagues, people are more confident when opinions are in agreement as opposed to disagreement. Although agreement is a signal of accuracy, it is also a signal of a shared perspective and shared error. People rarely recognize this latter implication of agreement. In fact, Ilan Yaniv, Shoham Choshen-Hillel, and Maxim Milyavsky have shown that confidence increases even when people understand that others’ opinions were cherry-picked to agree with their own initial answer.

To tap into the crowd’s wisdom, appreciating the roles of both knowledge and diversity are essential. People value the knowledge of individuals, and they often chase the expert to obtain it. But in doing so they may forsake diversity and risk missing out on the combined knowledge of the collective.

Jack B. Soll, Albert E. Mannes, and Richard P. Larrick

**Further Readings**


**Word Learning**

Language learning raises unique problems of learning and memory. This is widely recognized with respect to syntax learning, but it is also true of word learning. Word learning is the process of developing generalized (i.e., abstracted) mental representations to associate a *word form* (e.g., sequence of speech sounds or hand shapes/movements) with a *meaning* (e.g., category of events or objects that the word refers to) and *conditions of use* (e.g., Where in a sentence does this word typically belong? In what social contexts does one use the word?). The remainder of this section describes some unique features and questions about word learning in comparison to other kinds of learning. The next section describes research findings on children’s word learning, including the typical course of vocabulary development, individual differences, typical errors, and ecological and cognitive factors that facilitate word learning. Subsequent sections briefly describe the neurological changes associated with word learning, the relation of word learning to reading, the nature of word learning in multilingual individuals, and word learning in adulthood.

Word learning entails special questions because the corpus of words we learn, our *lexicon*, is a unique set of information. It is dynamic and additive: Consider how the compound word *electronic*
mail, coined in the 1980s, was quickly reduced to e-mail, which has since spawned analogous terms such as e-commerce. Adults can rapidly understand such words despite their novelty. This illustrates how we can, throughout life, add new elements (words) to our lexicon. In so doing we establish new, systematic connections (of sound, meaning, syntax, and usage) to other words and other linguistic and conceptual knowledge. Although words are arbitrary in form (e.g., nothing about the sound dog is inherently doglike), the lexicon is nonetheless somewhat principled. For example, words are hierarchical in meaning (e.g., animal refers to a category that includes all referents of dog) and in structure (e.g., an -ed verb ending denotes the real or hypothetical completion of an act or state). Also, words are combined in particular ways to express more precise meanings (e.g., fire truck and truck fire have different meanings). The lexicon is both social and normative (e.g., only our cultural knowledge makes e-commerce understandable) and internalized (e.g., we use words to facilitate cognitive processes such as explicit memory).

Word learning can be called symbol learning because it encompasses not only spoken words but signed words and even pictorial symbols (e.g., brand logos). Several nonhuman species (i.e., apes, parrots, dolphins) can learn small numbers of abstract names and symbols for objects, properties, or actions. There is no evidence that nonhuman animals use the full human range of word meanings (e.g., not, think, silly, maybe), word variants (go, gone, went, had gone), or word functions (e.g., puns, metaphors, novel compounds such as “climbing wall”). Yet children as young as 2 to 4 years old flexibly adopt such a wide range of forms, meanings, and uses: They can learn words defined by tone variations (e.g., Mandarin; Yoruba), percussive “click” or ingestive noises (Sindhi, Xhosa, Zulu), or gestures (American Sign Language). They learn words that take complex inflections (i.e., changes to the forms of a word, such as run, ran, running). Such variations are extensive and complex in languages like Turkish and Hungarian. Children also can integrate word meanings with cultural and conceptual knowledge (e.g., American children know that Pokémon refers to fictional characters, toys, playing cards, a game, a TV program, DVDs, and a video game, but Pokémon Diamond only refers to the last of these). How do children learn all of this?

How Children Learn Words

Vocabulary Development in Childhood

The course of word learning in young children is somewhat predictable. The first 50 to 75 words are acquired slowly, typically by 18 to 24 months of age. These words include proper names (Mama), nouns (bottle), a few verbs (give), descriptive terms (down, more), and social routine words (bye). Words like nothing or think are absent. The largest subset of early words is nouns, though it has been argued that this is not true of all languages.

Although first words are acquired slowly, even infants recognize a few words: Their own name sounds familiar by 4 months and by 10 to 12 months they tend to look selectively at an object when they hear its name (“truck!”). Many infants say a few words around 12 months (infants who are learning signed languages gesture their first words around 9 months). Then around 18 to 24 months, toddlers start learning words faster. In Indo-European languages (e.g., English, Italian, Dutch), infants start learning nouns faster, until their total receptive vocabulary (i.e., the words they understand) includes 150 to 200 items. Subsequently, the proportion of new verbs and adjectives increases relatively faster. This suggests that children learn nouns until they can and must express more diverse and specific relations between nouns (e.g., “The man petted the dog,” vs. “The man fed the dog”). This requires verbs. Fittingly, toddlers start producing two- and three-word sentences around 18 to 24 months. These protosentences are telegraphic: They lack articles, prepositions, and inflections. Only after children understand 200 to 400 words (2½–3 years) do they add many such function morphemes to their sentences.

From 3 to 5 years of age, vocabulary grows substantially. Although total vocabulary size becomes harder to measure, English-speaking first graders might know an average of 3,000 root words (i.e., uninflected terms such as house, run), and many more inflected or compound words (e.g., running, houseboat). These large gains have spurred folk beliefs that children are uncannily precocious word learners. However, this claim lacks specificity or verification, and adults in controlled tests learn new words faster than preschoolers. Thus, acceleration in word learning around 18 to 24 months and large vocabulary gains from 2 to 5 years do not prove
that word learning is a specialized childhood learning ability.

**Individual Differences in Children’s Vocabulary**

Throughout childhood there is great variability in individual vocabulary size. According to parental report data, average English-learning 24-month-olds use about 300 words. However, children in the lowest 10% use only about 50, and those in the highest 10% use about 500: a 10-fold difference. Adults have similar large differences in vocabulary. At the lower extremes, virtually all children with cognitive or language disorders have some sort of restricted vocabulary.

Among children with language impairment but no other cognitive deficits, a common problem is that the auditory system (i.e., brain network that processes sound information) is slow to process the sound information in speech. This will inevitably impair word learning because, for example, it is harder to separate individual words in continuous speech. It now seems that this problem leads to later problems in decoding words while reading.

The Setting for Children’s Word Learning

How do children learn words, and what factors influence children’s success in word learning? As to “how,” the simplest answer could be that children hear words while attending to the referent and form an association between the two. However, this explanation is inadequate. There are so many possible associations that a more specific theory of learning processes is necessary. A traditional associationist account holds that learning requires words and referents to be paired (a) close in time and (b) frequently. Both assumptions are only partly supported. First, in some situations, toddlers associate a novel word with something they saw a few minutes ago, not the last thing they saw. Second, frequency of input does not precisely predict learning. Preschool children sometimes sensibly guess a word’s meaning from hearing it only one to two times. Even infants, after hearing a new word only a few times, might remember something of that word’s sound for days. However, such fast mapping has been documented in simple, unambiguous experimental contexts, where adults use ostension (i.e., naming while showing the child a referent). Ostension is used by parents in specific situations, like picture-book reading. It is unknown how much fast mapping happens in common, everyday situations. Even in moderately complex experimental tasks, young children require many repetitions to learn a word. A correlation has been found between how much parents speak to infants (i.e., variety of words and total words) and the infants’ vocabularies several years later. Thus, although repetition is not all determining, it usually promotes learning.

Children learn words even when adults are not providing ostensive naming or speaking to them at all. In many cultures, adults speak to infants infrequently or not at all. Perhaps surprisingly, there is no evidence that those infants learn language slower or have smaller lexicons than infants who are spoken to. Thus, the correlation between amount of speech to infants and later vocabulary does not rely on direct speech to infants. Infants must learn a lot by overhearing other people talking. Experiments show that toddlers may learn words as effectively from overhearing as from direct ostension.

The Progress of Word Learning: Errors

Children’s knowledge of a word does not simply fit into one of two binary states, learned or unlearned. Children, like adults, can know a little about a word (e.g., “It sounds familiar . . . maybe it’s a kind of food . . .”) or a lot (e.g., can recognize, define, and use it correctly), or anything in between. However, children show a lag from comprehension to production. That is, they typically understand a word before they will say it. This is partly because of slow development of the fine motor skills for speech production. Nonetheless, children do speak, and this can reveal what they know or do not know about a word’s meaning. Children’s characteristic errors include overextending words (e.g., calling any medium-sized mammal “kitty”). These errors sometimes reflect real confusion about a word’s meaning (e.g., kitty = any cute, fuzzy pet) and sometimes reflect pragmatic accommodations to their small vocabulary. If your only animal words are kitty, horsey, and birdy, your best option for labeling a rabbit or squirrel is “kitty.”

How do children correct errors like these? Occasionally children seek information (e.g., “What dat?”), but children often do not seem to realize they are making errors. However, parents sometimes correct children’s overt errors of meaning or
word choice. They also use less direct strategies, like expanding and elaborating on their toddlers’ telegraphic statements. A 20-month-old might point to a pond and say “duck!” The parent might then expand, “Yes, the duck is swimming, isn’t it?” This expansion might teach the child not only about swimming but also confirm the correct usage of duck and implicate a semantic relation between duck and swim (i.e., ducks are a sort of thing that swims). Or if the parent elaborates, “Yes, ducks are pretty birds!” this provides semantic information about the class-inclusion relation between the categories duck and bird. Regarding the individual differences noted above, parents who speak more to their infants (who will later have larger vocabularies) also tend to elaborate. Parents’ expansions might therefore provide important input to toddlers about word forms, meanings, and uses.

Children’s Readiness to Interpret Words

Young children are not “equal opportunity” learners, assigning any plausible meaning to a new word. Children have certain biases. Some are based on perceptual processes. In general, objects that are novel, bright, and prominent will be associated with a novel word. Also, infants tend to associate a new word with an object if the object is moved in rhythm and synchrony with repetitions of the word. Finally, children tend to map novel words for objects onto whole objects, as opposed to specific parts, colors, or textures. However, more specific information about the word can cause children to override their bias and associate the word with another property.

Other biases in interpreting words seem to reflect human conceptual knowledge and ignorance. For example, children seem to assume that words refer to categories of objects, events, or properties, rather than to individuals. A child hearing “lemur” will associate it with a category of similar animals. Although some of toddlers’ first words may be narrowly context specific, this seems to be the exception rather than the rule. Even by their first birthdays, infants tend to generalize new words to classes of similar referents. Also, children, like adults, tend to generalize words for objects at a basic level of abstraction; that is, an intermediate-breadth category (e.g., car) rather than a very specific one (Mazda 626) or a very general one (vehicle). More specifically, once toddlers know 50 to 100 words they begin to assume that novel object words generalize to categories of same-shape objects. However, this is a learned bias, and it is contingent on other properties (e.g., is the object an animal or artifact?). Thus, conceptual biases are not freestanding: They rest on other experiences and learned patterns. For this reason, it is no surprise that children’s language constrains the specific concepts that they learn and name. Cross-linguistic studies confirm that meaning biases are affected by language experience. For example, English and Korean prepositions denote different spatial relations: English in and on do not have exact analogs in Korean. Korean 1-year-olds are sensitive to spatial relations denoted by different Korean words, but English-learning 1-year-olds do not discriminate those relations. Thus, toddlers’ lexicons influence their sensitivity to specific meanings and patterns in the world.

Children also have social biases that affect how they learn meaning. By 18 months of age, toddlers monitor where adults are looking, so that when the adult says a novel word, the toddler associates it with whatever referent the adult was looking at. This prevents the infants from spuriously associating words with whatever they are attending to, if the toddler and adults are attending to different things.

Other biases for inferring word meanings are ambiguous. One claim is that children believe that each nameable category only has one label—a mutual exclusivity bias. For example, if the child sees a horse and a tapir, they will assume that an unfamiliar word (tapir) refers to the unfamiliar animal. However, evidence does not support that this is children’s true bias. They do preferentially associate a novel word with a novel rather than familiar referent, but there are many possible explanations for this. This exemplifies a general pattern: Although children have many biases for interpreting new words, it is not clear which, if any, of these biases are specific to word learning per se.

Words on the Brain

To understand word learning requires understanding how sound patterns of words are processed by the brain and represented by brain networks so that subsequent brain states (caused by, e.g., the sound of the word) will reactivate that word representation. Activation of lexical knowledge involves widely distributed networks in the cortex, but in most healthy
adults it persistently (not exclusively) involves left frontal and temporal cortical areas. However, this anatomical specialization is the result of development: Infants show wider distributed and more bilateral activation during word processing. Activation becomes more focused in left temporal and parietal regions from 14 to 20 months, showing that neuroanatomical specialization starts early. Intriguingly, infants who understand at least 150 words show a more focused electrophysiological response to familiar words as early as 200 to 400 milliseconds (ms) after the word begins.

Word Learning and Reading

As children get older, they can decontextualize language—see it as separable from the “here and now” (e.g., talk about absent referents, tell stories). Decontextualization of language is maximized in written text, such that we can enjoy the language of “speakers” who are absent, or even deceased.

Learning to read during childhood has a bidirectional relation to word learning: Children with larger vocabularies do better in reading, and children who read a lot learn more words. Throughout school, vocabulary is the best predictor of reading comprehension. During grade school, some nontrivial proportion of vocabulary growth is due to word learning from text. When unfamiliar words arise, we try to use the meaningful content of surrounding text to interpret them. Although a minority of contextually learned words are retained, the consequences are nevertheless substantial: Hypothetically, if a child reads 600,000 words in a summer (e.g., the last three Harry Potter books) and 1% of words are unfamiliar but inferable from context, and if she has only a 5% chance to infer and remember a word from context, her net gain would be 300 words. Thus, reading a lot of grade-level text is important for vocabulary growth. Children at risk for reading failure enter school with a lower level of language skills, read less, and remain below-average readers with smaller vocabularies.

In skilled readers, recognition of written words elicits maximal activation in a specific region in the temporal cortex. Less skilled readers show too widely distributed patterns of activation over many cortical regions, and recent evidence suggests that training these readers’ discrimination of sound patterns in words can lead to more focused patterns of brain activation during reading.

Words in Two Languages

Most people in the world are multilingual: Monolingualism is the exception. How do people learn two lexicons, which might overlap in meaning but contain many single-language word forms? One debate is whether two lexicons are initially merged or separate. Although there is great diversity across individuals and situations, toddlers’ two languages begin separating very early. Recent brain research suggests that bilinguals show activation of largely but not completely overlapping areas of cerebral cortex for each language.

Word Learning Later in Life

Word learning continues throughout life. There is a general idea that age of acquisition matters: Words learned earlier (e.g., as a toddler) are the most strongly represented in neural networks. For example, in aphasia, or loss of language due to brain injury, there is usually some degree of anomia (i.e., poor production or understanding of words). However, early learned words are more likely to be retained.

Word learning in adulthood can be very robust. Some words learned as a young adult will be retained for decades, even if never heard or used in the interim.

What do we know about the processes of word learning in adults? Like children, adults learn most new words by inferring meaning from context. Adults’ rich phonological knowledge helps them efficiently learn new sounds of words. Adult word learning is affected by many general cognitive effects: For example, words at the beginnings or ends of sentences are more likely to be remembered (i.e., primacy and recency effects: the general advantage in remembering items from, respectively, the beginning and end of a list). When learning words over time, distributed practice rather than massed practice tends to increase retention intervals (i.e., how long words are remembered). Associations of new words are subject to both proactive and retroactive interference (i.e., confusion caused by prior information or subsequent information, respectively). In all these effects, we see continuity from childhood to adulthood, and substantial overlap of word learning with general processes of learning and memory.

Gedeon Deák
See also Concepts, Development of; Dyslexia, Developmental; Language Development; Language Development, Overregulation in; Representations, Development of; Speech Perception; Word Recognition, Auditory; Word Recognition, Visual

Further Readings


**Word Recognition, Auditory**

Language provides humans with the remarkable capacity to express their thoughts through a physical medium to share with others. To do so, we combine elements, words, whose form has been conventionalized within a particular language community. Thus a critical step in the process of retrieving a talker’s message consists of identifying these elements in his or her speech. This entry discusses how our knowledge of the auditory forms that words take may be represented in memory, and how listeners decide, based on the auditory stimulus, which words they heard, out of all possible word combinations the talker may have spoken.

**What Does Our Knowledge of Words Look Like?**

When we listen to someone talk, words seem to pop out of his or her speech effortlessly. This impression is misleading, however. Words are not neatly segregated from one another in speech as they are in print. How many words the utterance contains, and where they begin and end in the speech stream, are properties that the listener must establish. Moreover, the way spoken words sound varies considerably across contexts—for example, when produced by a man or a woman, in the clear speech used in lecture halls, or in the casual speech characteristic of informal conversation. Our knowledge of the form of words must accommodate this variability. Two approaches to this issue can be contrasted.

First, listeners may represent the form of a word as a compilation of the memory traces that correspond to all past exposure with the word. Each instance retains the acoustic properties resulting from the context in which the word was uttered. Such a representation is sometimes described as a cluster of observations in a multidimensional space. A more compact representation may also be postulated, such as one that represents the central
tendency derived from past instances of a word, its prototype. These views assume ever-changing word representations because new instances of words are constantly added to the cluster or the set of instances that contribute to the central tendency.

These exemplar and prototype views are rooted in cognitive theories of categorization and contrast with a second, linguistically grounded, approach where words are represented by the features that distinguish them from other words. The acoustic properties of a spoken word, such as the voice quality of the talker that utters it, are considered irrelevant to this distinction and consequently not part of the representation of the word's form. This approach assumes abstract, context-independent, and immutable representations. Normalization algorithms transform information extracted from the speech to neutralize the influence of contextual variability, in effect treating it as noise, or to model the variation and factor out its influence.

Distinguishing between the two approaches has proven difficult. For instance, some have taken the fact that people recognize words uttered by familiar talkers more readily than the same words from unfamiliar talkers as evidence supporting the instance-based approach because it demonstrates that nondistinctive properties of spoken stimuli are maintained in memory and contribute to recognition. However, the finding is also compatible with the abstractionist approach if one assumes that the normalization algorithms can be optimized to reflect past experience with a given talker.

**Does Recognizing a Word Require Recognizing Its Parts First?**

Another widely discussed issue concerns the internal structure of words. Phonological theories describe words as built out of elements, the phonemes, grouped into larger units, such as syllables. Words' internal structure is known (albeit implicitly) to language users because changes that word forms undergo under the influence of morphology or other linguistic constraints have regularities that depend on the decomposition of word forms into such a structure. The critical question here is whether people, when analyzing speech, decompose the signal into individual elements to establish which word matches the structure best. Recognizing phonemes or other units first, as opposed to analyzing the spoken word as a whole, may offer an advantage because there are fewer phonemes to discriminate than there are words. However, phoneme recognition itself has proven difficult because the acoustic realization of a given phoneme varies greatly across contexts. Furthermore, listeners can successfully retrieve which of the phonemically identical strings (such as *two lips* and *tulips*) the talker said because they make use of subtle acoustic differences between the strings. This finding is difficult to explain if the speech signal was first translated into its phonemic subcomponents.

**Word Recognition as a Perceptual Choice**

In contrast to the question of how the form of words is represented, the process by which the perceptual stimulus is compared to these representations is relatively well understood and uncontroversial. Spoken words become available to the listener over time. Because speech is a complex, transient, and rapidly changing signal, and, because sensory memory is limited, speech must be evaluated and interpreted incrementally rather than word by word. But the early portion of a spoken word (e.g., *cap ...*) is often compatible with many different words (e.g., *cap, captive, capital, captain*). Dominant views posit that all possible interpretations of the spoken word can be simultaneously considered. For example, in William Marslen-Wilson’s cohort theory, the first sounds of a spoken word determine a cohort of hypotheses compatible with this early information. Subsequent information serves to prune the hypotheses no longer supported by the signal. Although the “propose-then-dispose” aspect of the theory has since been falsified by evidence that words can be successfully recognized even when their first sounds are distorted, the privileged role of the early portion of a word has been maintained by assuming a form of competition among simultaneously considered hypotheses. The more evidence has accumulated in favor of a given hypothesis, the less likely its alternatives. This mechanism, in effect, favors words that match the early portion of the spoken stimulus over those that match a later portion, because the latter will have been largely discounted before the stimulus supports them as possible contenders. Importantly, competition is modulated by the likelihood of encountering each hypothesis, which can be estimated by how often it has been encountered before. Frequent words are recognized more accurately and faster than rarer words, and frequent hypotheses interfere with the recognition of rare words.

*Delphine Dahan*
Word Recognition, Visual

Reading is one of the most remarkable of our cognitive abilities. Skilled readers are able to recognize printed words and compute their associated meanings with astonishing speed and with a great deal of accuracy. This level of performance arises despite the fact that letters frequently appear in an unfamiliar form (e.g., in new fonts) and constitute a limited array that renders individual words highly confusable (e.g., salt, slat).

This entry provides an overview of some of the key theoretical claims about the cognitive architectures and processing mechanisms that underlie visual word recognition. These claims were first instantiated in the interactive activation model developed by James McClelland and David Rumelhart in 1981, from which many of the more recent theories in the field have been developed. They are supported by evidence from a variety of experimental methods, including observation of word recognition performance in skilled readers (e.g., measuring the time taken to read a word aloud), investigation of the reading behavior of people with acquired or developmental language impairments (e.g., dyslexia, pure alexia), and computational modeling (e.g., testing theories of visual word recognition through computer simulations of human performance).

The Architecture of the Visual Word Recognition System

Though the earliest theories of visual word recognition claimed that words are recognized as wholes on the basis of their shapes, modern theories suggest that words are recognized in a hierarchical manner on the basis of their components. Information from the printed stimulus maps onto stored knowledge about the visual features that make up letters (e.g., horizontal bar, left-opening curve), and information from this level then proceeds onto a system of stored abstract letter representations that code letter identity as well as letter position (so that anagrams like top, pot, and opt can be distinguished). These letter representations are abstract in the sense that they can be activated irrespective of surface characteristics such as case, size, font, and retinal location. Information at the letter level of representation then proceeds onto an orthographic lexicon (a body of stored knowledge about the written forms of whole words). Units in the orthographic lexicon can then activate information about the meanings and/or sounds of words. Visual word recognition is thought to be achieved when a unit in the orthographic lexicon reaches some critical threshold of activation.

There is widespread agreement that each unit in the orthographic lexicon is coded in terms of an individual’s experience with that word. Precisely how lexical experience is best conceptualized is a matter of some debate, however. Until recently, most theories argued that orthographic units are coded in terms of the frequency with which a word occurs in the language, and indeed, word frequency is known to be the most powerful determinant of the time taken to recognize a word (i.e., its latency). However, recent research has suggested that the age at which words are acquired, or perhaps the cumulative frequency with which an individual encounters words over his or her lifetime, may provide a better means of conceptualizing lexical experience. Both age of acquisition and cumulative frequency have also been shown to influence word recognition latencies, though because age of acquisition, cumulative frequency, and word frequency are naturally correlated, it is not yet known which variable provides the optimal index.

Processing Dynamics in Visual Word Recognition

Information is thought to flow through feature, letter, and whole-word orthographic levels of representation in an interactive manner, such that information at higher levels of representation can influence processing at lower levels of representation. The finding
that provided the initial support for interactive processing is the word superiority effect. Participants are better able to decide which of two letters (e.g., D or K) is in a briefly presented target masked by another stimulus (e.g., immediately followed by hash marks) when that target is a word (e.g., WORK) than when it is a nonword (e.g., OWRK). This finding supports the notion of interactive processing because it suggests that a decision based on activation at the letter level is influenced by higher level information from the orthographic lexicon. More recent research has shown that tests of visual word recognition such as speeded lexical decision (i.e., deciding as quickly as possible whether a stimulus is a known word) show top-down influences of semantic variables such as imageability, number of semantic features, and number of meanings. Similarly, substantial evidence also suggests that the recognition of printed letter strings is influenced at its earliest stages by information about the sounds of words.

It is generally thought that printed letter strings activate multiple candidates in the orthographic lexicon (e.g., the stimulus cat activates units for cat, cab, rat, mat, car, cut, etc.). The activation of multiple units thus raises the question of how the target unit is ultimately selected. Though theories are divided on this issue, one popular mechanism is competition. Inhibitory connections between units in the orthographic lexicon enable the most active unit (usually the one corresponding to the target) to drive down activation of multiple alternative candidates. One key piece of evidence for competitive processing is that presentation of a high-frequency, masked stimulus (e.g., able) tends to inhibit recognition of an orthographically related target (e.g., axle). There is also mounting evidence for a processing cost involved in the recognition of words with higher frequency orthographic neighbors (i.e., orthographically similar words), though further research is needed to establish this finding conclusively.

Kathleen Rastle

See also Compound Words, Processing of; Dyslexia, Acquired; Eye Movements During Reading; Frequency Effects in Word Recognition; Word Recognition, Auditory

Further Readings

**Working Memory**

*Working memory* is the term used to describe the information one is thinking about at any particular moment. That information keeps changing, and the amount one holds in mind in this way at any moment is quite small compared to the vast amount of information in one’s permanent memory storage system in the brain. The concept of working memory has become one of the most important and often-used concepts in the field of psychology, as it helps a great deal in explaining what tasks are easy or hard for individuals to carry out. There are consequently thousands of articles on various aspects of working memory.

The term working memory was suggested in 1960 in a book by George Miller and his colleagues, and the concept was made popular 14 years later in a book chapter by Alan Baddeley and Graham Hitch. The idea behind the term is that there are various kinds of mental work such as thinking, problem solving, reasoning, language comprehension and production, and keeping track of changing events (i.e., while watching a baseball game). To do these kinds of mental work, one must hold in mind certain information relevant to the situation. One might need to hold in mind data such as, in baseball, which team is at bat and how many outs there are in the inning or, when doing addition in one’s head, the carried digits. One might also need to hold in mind plans such as, when solving an arithmetic problem, the steps to follow or, when running errands, the order in which errands are to be accomplished. The key point is that the amount of information that can be held in working memory is limited. This fact in turn puts limits on how well humans can solve problems, formulate plans in their heads, and so on.
Working memory appears to have various components, but they operate together as an integrated system. For example, there appears to be cross talk such that, if the goals are too complex, one can forget not only goals but data; conversely, if there are many data to be kept in mind, one can forget not only data, but goals.

This entry will include a discussion of many important aspects of working memory, including a comparison with similar concepts and terms, the kinds of studies demonstrating working memory, limits on working memory, theories of working memory, individual and group differences in working memory, and some possible neural and evolutionary reasons why working memory capacity is limited.

**Similar Concepts and Terms**

One of the most common questions asked of researchers of working memory is how it differs from certain other types of memory. Usually, there is a lot of overlap between the different types of memory in question. Here are some of these overlapping concepts and how they subtly differ from the term working memory. This discussion is meant to resolve some of the confusion that inevitably comes about when so many terms are used.

**Limited-Capacity System**

In some of the early work on working memory, a single term was sometimes used for many different types of things that were limited in the mind. We can only attend to a limited number of objects in the visual field at once. We can only keep in mind a small number of randomly arranged letters, numbers, or words at once. We can only fully comprehend one talker at a time. We can only solve math problems in our heads if they are sufficiently simple. A term often used to describe these limits is a limited-capacity system. Capacity is the ability to carry out tasks, and this capacity is limited in various ways just described.

Are all these limitations based on the same, single, limited-capacity part of the human mind and brain? It may be, but it would be difficult to prove, and no such proof has emerged as of yet. There are shortcomings of that way of thinking. For example, your memory for just how the final note of a symphony sounded, just after the end of the symphony, may linger at most a few seconds, without interfering with your ability to carry out a math problem. If you finish the math problem and then turn attention to the sound that has just finished, you may still be able to experience that sound through memory (though you will have missed much of the immediately preceding part of the symphony). If the memory for sound and the math problem activity are indeed separate from one another in your mental processes then they should not both be considered part of a common limited-capacity system.

The math problem is the kind of activity more often considered dependent on a limited-capacity system; it requires attentive thought. Passively remembering a few tones, words, or images, on the other hand, is generally considered to be partly outside of that limited-capacity system because some such memory continues for a short time even when your attention is elsewhere. Yet both the active, attentive and the passive, inattentive types of mental processing typically contribute to performance on tasks considered to be working memory tasks.

**Immediate Memory**

Immediate memory is memory for a list or array of items that was just presented. There is no delay between the stimulus and the time at which you are supposed to try to recall that stimulus, so it is called immediate. Tasks using immediate memory procedures are probably the most common types of task used to study working memory.

**Short-Term Memory**

This term is used in different ways by different folks. When many people say “I’m having trouble with my short-term memory” they are saying that they cannot recall things that they did within the last day or so, such as where they parked their car in the morning. That kind of memory clearly cannot be the same as the information currently held in mind, unless one spends the entire day trying to bear in mind where the car was parked.

When psychologists use the term short-term memory, they are contrasting it with the term long-term memory, the vast amount of information that we have learned over a lifetime. In this sense, short-term memory is the same thing as working memory. One difference is that some researchers use the term short-term memory to describe only the passive, effortless storage of information for a few seconds. Some of them reserve the term working memory...
to refer only to temporarily held information that does require attention, whereas others use the term working memory to refer to all the temporarily held information—both active and passive. For the latter researchers, short-term memory is a part of working memory. (Differences in how researchers use the terms are regrettable, but they are hard to control. Researchers would like to be able to communicate with one another and with other people in a standard manner, but slightly different interpretations of the term are already embedded in a lot of published literature.)

**Sensory Memory**

The memory for exactly how a sound sounds, how a visual scene looks, how a caress feels, or how a food item smells or tastes is sometimes impossible to put into words. Research shows that our memory for sensory qualities is excellent and can include myriad things in the environment at once. This kind of memory, however, fades within a few seconds. In that time we are able to concentrate on a few items and save them in a more categorical form. In one type of experiment, for example, an array of characters is very briefly presented; perhaps three rows of four randomly chosen letters are presented. Then the array disappears and a tone is presented to indicate whether the top, middle, or bottom row is to be recalled. If the tone comes very soon after the character array, almost all the letters in that row can be recalled. If the tone is delayed, the ability falls off. We can retain only about four of the characters but, if given prompt notice by the tone, we can choose which four characters to draw from a rich but short-lived visual sensory memory and to encode in a verbal form. Visual sensory memories are richer in spatial details, whereas auditory sensory memories are richer in precise details about the timing of the sounds.

Most researchers do not consider sensory memory as part of working memory, but it cannot be denied that sensory memory plays an important role in tasks used to measure working memory. For example, the memory for a spoken list of words is typically better than the memory for the same words presented visually, and this auditory superiority comes from items at the end of the list, where auditory sensory memory preserves vowel sounds in the last word or so and makes these words easier to remember.

**Activated Memory**

One way to think of working memory is that it consists of the elements of memory that are currently in a heightened state of activation. For example, whereas long-term memories might be preserved in terms of the pattern of synapses between nerve cells that have developed over time, working memory might be represented by the temporary activation of some patterns of neural activity that represent the concepts currently in working memory. Of course, we often perceive things that are different from anything we have perceived before, in which case the information has to be added to long-term memory while it is being activated. For example, have you ever thought of carrots that are purple? That thought has just been added to your long-term memory, and it is also currently active, which may make up your working memory of this novel variety of vegetable.

**The Focus of Attention**

Often, it is emphasized that attention makes a very important difference in memory. This certainly seems to be true in the field of working memory. Items to which one pays attention can be understood much more completely than items outside one’s attention (for example, what people in the room you are in are saying to one another while you are engrossed in a telephone conversation). If you want to form new, strong connections between items or ideas you must pay attention to the ideas. For instance, you cannot memorize directions from a map without intensely attending to the map.

**Studies Demonstrating Working Memory**

Even though we use working memory to help carry out a variety of problems, researchers often want to measure working memory in a manner as simple as possible so that the measurement will be valid for a range of types of problem to which working memory can be applied. Tests of intelligence have long included tests of *digit span* in which series of digits are presented and span is defined as the longest list that the participant can repeat back without error. Typically, young adults can repeat lists of about six or seven random digits. Other times, lists of letters or words are used and the results are similar. One can present series of shapes that cannot easily be named and the results are similar. People seem to do best at the beginning of the list (called the primacy
effect). If they are allowed to recall the items in any order, they do very well also for items at the end of the list (called the recency effect).

Individuals tend to use strategies to remember lists as well as possible, such as trying to group words together or repeating them silently (which is called covert verbal rehearsal). Such strategies can be prevented by requiring that the individual recite a meaningless word (such as the, the, the) while the items to be recalled are presented. This typically interferes most with memory for items at the beginning of the list. Some researchers are most interested in the best memory, strategies and all, but other researchers are more interested in understanding how much can be retained even without the use of such strategies.

Instead of lists, other studies have used arrays of characters presented all at once on the computer screen. The results are similar, although rehearsal may be more difficult for arrays than for lists.

Some studies are designed to find out how well people are able to save information while it is being used. The studies may present lists that require both saving information and using it, at the same time. These procedures will not necessarily look the same as tasks in the real world (such as problem solving or language comprehension), because in the real world it is not easy to find out how much the task depends on working memory and how much it depends on other abilities. In some working memory tests, therefore, the storage and processing parts of the task are interleaved rather than integrated. The research participant may have to answer a math problem, memorize a word, answer another math problem, memorize a second word, and so on. The number of words that can be remembered and repeated back, despite the interleaved math, is known as operation span.

Limits on Working Memory

Chunk Limits

Humans are, of course, impressive thinkers who often come up with ways around their limitations. It appears that the main limit on the ability to remember information is how many meaningful units must be remembered. These meaningful units are called chunks. Most people would have great difficulty in committing to memory a nine-letter series such as BIMICASUA. It is much easier, however, to remember the series IBMCIABFI, if one notices that it is composed of three chunks, each a meaningful acronym: IBM, CIA, and FBI. Some individuals have learned to repeat back lists of up to 80 digits, but they do so by having already memorized many series of digits that make up athletic records (like a running time for the mile of 3.86 minutes) and using that information as a chunk when that sequence happens to come up in the list. You probably cannot recall such long strings of digits, but everyone uses chunking a great deal. You could probably recall many, many words if they make up the words of the first verse of Mary Had a Little Lamb followed by the first verse of Twinkle, Twinkle, Little Star.

A great deal of research—for example, that summarized by Nelson Cowan and his colleagues—shows that adults generally can remember only three of four chunks of information. Sometimes each chunk includes a lot of information, but it is held by a very limited working memory mechanism. This limit may have to do with how much information can be attended at one time.

Time Limits

Baddeley and Hitch showed that chunks are not the only things important for retaining information in working memory. They found that lists of long words cannot be remembered as well as lists of the same number of short words, even if the long words are equally familiar. This difference has to do with the ability to use covert verbal rehearsal to keep words active in memory. It takes longer to rehearse long words and, while some words are being rehearsed, other words in the list can be forgotten.

Pierre Barrouillet and his colleagues have found that what is important for retaining items in working memory is the ratio of free time to time occupied by distractions. The more free time, the better.

Theories of Working Memory

The theories of working memory are theories of what can and cannot be remembered and why. If we had a complete understanding of what can and cannot be remembered, we probably would know
which theory is correct. There are many subtleties, though, and researchers still are investigating this issue.

**Central Storage Theory**

According to the central storage theory, working memory requires attention (a process that is central in the mind, not specific to one type of stimulus or another). If you add one working memory task, it will interfere with other working memory tasks and with other kinds of effortful thinking. There is some truth to this inasmuch as simply having to remember a series of about seven digits—what is called a memory load—can impair one’s concurrent ability to reason out problems logically and make good decisions. (Similar to memory load, one reason that too much alcohol results in bad decisions is that it impairs working memory.)

J. Scott Saults and Cowan carried out research on memory for both visual and auditory memory arrays together (colored spots on the screen and spoken digits in different voices from four loudspeakers at once). They found that participants could remember a maximum of about four items. If they were allowed to pay attention only to the colored spots, they were able remember about four of those. If they were required to pay attention to both modalities, they were able to remember still about four items total: about three colored spots and one spoken digit. This suggests that a central storage mechanism exists, though it certainly does not show that this central storage is the only kind. For one thing, sensory memory had to be eliminated using a meaningless audiovisual pattern, or mask, after the arrays in order for this fixed capacity of about four items to be observed.

**Modular-Stores Theory**

Some other theorists, such as Robert Logie, believe that there is no central memory, only separate memories for items stored in visual form and items stored in verbal form, examples of specialized subsystems or modules. These theorists attempt to explain results such as the ones using visual and auditory stimuli together by saying that items are thought of in a different way or recoded. One can make up a verbal code for a visual stimulus (like naming a colored spot), or one can make up a visual code to go with an acoustic stimulus (like envisioning how a spoken digit would look in print). We know that stimuli that are similar in their codes interfere with one another in working memory. For example, it is difficult to remember the printed letter series c, d, u, p, t, z, b because the letters tend to be mentally recoded into a speech form and they sound similar. This shows how difficult it will be to determine with great certainty whether there is a central store.

**Hybrid Theories**

Today many researchers believe some sort of hybrid model that includes both (a) a central store that is closely tied to attention and accepts a variety of types of information and (b) some other forms of storage that are more passive and automatic and may be designed to accept only certain specific kinds of information. For example, for attention-related memory, Cowan thinks of the focus of attention as a central storage device in the mind, and Baddeley talks of an episodic buffer that contains links between information of different sorts (like an association between a face and a spoken name). In both theoretical views, there is also a central executive component that represents the attention-driven control of information as it is transferred from one state or store to another but that presumably does not itself store information.

For passive storage devices, Cowan talks of various sorts of activated elements from long-term memory, and Baddeley talks of visual and phonological storage buffers as separate modules. One difference between these models is that Cowan is unsure whether the types of activated long-term memory are few enough or simple enough to be considered modules, or whether there are instead just a myriad of different kinds of activated memory for different kinds of features of the stimuli (color, spatial location, sound quality, semantic meaning, geographic arrangement, and so on).

**Individual and Group Differences in Working Memory**

Individuals with lower scores on working memory tests have been shown to do worse on many different kinds of cognitive tests. They remember fewer items in working memory, find it more difficult to inhibit irrelevant thoughts, and find it more difficult to remember the goal of an activity than do individuals with a larger working memory span. Therefore, working
Working Memory, Evolution of

memory is of great practical significance. Recent work has suggested that the attention-related part of working memory can be improved through training, and that it improves cognition and helps in the treatment of attention deficit disorder and dementia. If that information holds up, it will be exciting indeed.

Why Is There a Limit in Working Memory Capacity?

Various kinds of explanation have been given for the working memory limit. Saving information in working memory seems to depend on areas in the parietal lobes in the brain, and deciding what information to save and how to use it seems to depend on areas in the frontal lobes. The coding of a stimulus seems to depend on different areas that code different features all firing at once. For example, the representation of a blue circle would include neurons that represent blueness and neurons that represent circles firing at the same time. When several items have to be represented at once, there is the danger that they will corrupt one another. A blue circle and a green square that are parts of a set of items that is slightly too large for memory might get misremembered as a blue square and a green circle.

A few mathematically oriented researchers have argued that groups of three or four items make up an ideal grouping for the retrieval of information from memory. Items within a group of three are easily identifiable inasmuch as there is a beginning, middle, and end of each group. Evolutionary psychologists have argued that working memory is especially important in group interactions. One might have to figure out and bear in mind not only what a friend thinks but also what the friend thinks that he or she thinks, and what other people think that they both think. A good theory of the minds of others can help one be a leader in society. That takes working memory.

Nelson Cowan

See also Aging, Memory, and Information Processing Speed; Emotion and Working Memory; Intelligence and Working Memory; Rehearsal and Memory; Visual Working Memory; Working Memory, Evolution of

Further Readings


Working Memory, Evolution of

This entry discusses a prominent cognitive theory called working memory and discusses evidence for its evolution in the archaeological record. Working memory is a theory of cognitive function proposed by experimental psychologists Alan Baddeley and Graham Hitch in 1974. In their original formulation, working memory was conceived to be a multicomponent system that allows an organism to keep task-relevant information in active attention while filtering out task-irrelevant interference. At the core of this system was the central executive, a limited-capacity, attentional controller and decision maker. At its behest were two subsystems, the phonological loop and the visuospatial sketchpad. The
phonological loop consisted of two components: a quickly fading phonological store (about 2 seconds in length) and an articulatory rehearsal mechanism, which could be invoked vocally or subvocally. Repeated articulatory rehearsal of verbal stimuli was considered to have obligatory storage in declarative long-term memory. Baddeley and his colleagues found empirical support for the hypothesis that the phonological loop was critical to language comprehension and production. The visuospatial sketchpad maintained visual and spatial information in attention and played an important role in spatial orientation and wayfinding. More recently, Baddeley has proposed a fourth component, the episodic buffer. As originally conceived, the central executive had no storage capacity of its own; it was the analytical component of working memory. Because the phonological loop and visuospatial sketchpad consisted of modality specific, rapidly fading, limited-capacity stores, the model as originally conceived lacked a store that could hold integrated information. He therefore proposed a subsystem for the multimodal integration of verbal and visual information and the temporary maintenance of the resulting information for manipulation by the central executive.

Misconceptions About Working Memory

There is some confusion in the cognitive literature about the term working memory. Sometimes, the term working memory is used only in a narrow sense, and it does not imply Baddeley’s multi-component model. When it is used in the narrow sense, the term usually refers to the ability to maintain and manipulate thoughts over a brief period of time despite interference. Readers must often discover for themselves whether the term is being used in the narrow sense or the broad, multicomponent sense.

Evolution of Working Memory

Baddeley himself recognizes that evolutionary approaches are currently popular within psychology. His own investigations led him to believe in the selective value of the phonological loop as a powerful aid in the acquisition of language. Baddeley has also pondered the biological, adaptive functions of the other components of working memory. The two working memory subsystems (phonological loop and visuospatial sketchpad) would play a role in an organism’s perception of its environment. Baddeley recognized that because organisms must make decisions regarding their perceptions of the world, the central executive must play a prominent role in deciding on subsequent behavior based on information held in attention and analyzed. Baddeley noted that conditions of rapid change, such as those that would have confronted our early ancestors, would require the organism to be able to learn and learn quickly (i.e., implicit learning). Finally, Baddeley proposed that a successful system would be able to remember previous experiences and use this information in the creation of alternative scenarios (i.e., planning). Baddeley’s episodic buffer, described in his 2007 book, plays a central role in this planning function, and working memory would therefore reside at the intersection of cognitive functions, including perception, learning, attention, and action, which he believed enhanced the organism’s flexibility and survivability.

Recently, Frederick L. Coolidge and Thomas Wynn have proposed that an enhancement in working memory capacity was an important component in the long evolutionary trajectory of human cognition. Their model of cognitive evolution posits two major evolutionary leaps in cognition, one about 1.8 million years ago with the transition from sometimes living in trees (the Australopithecines) to fully terrestrial life (Homo erectus), and the second beginning somewhere between 100,000 to 30,000 years ago with the advent of fully modern behavior.

Several developments in cognition can be inferred from the archaeological and fossil evidence that marked the advent of Homo erectus. Among them are developments in spatial cognition, fine motor control (praxis), ability to imitate (perhaps driven by the motor neuron system), and a likely change in sleep. The transition from tree to ground sleeping aided in the integrity of a single sleep period, which may have released pressures against the deepest stages of sleep (slow-wave and REM sleep). This in turn may have led to a greater ability to consolidate procedural memories, a greater opportunity for threat rehearsal in dreams, and even creativity. One development in working memory is also evident. Homo erectus was able to coordinate visual information from the ventral stream (shape recognition) and dorsal stream (spatial arrangement) in the manufacture of stone tools. This ability was almost certainly deployed via the visuospatial sketchpad of working memory and indicates an advance in cognition over ape abilities.
Coolidge and Wynn’s second major leap in cognition was initiated by a transmissible genetic event that enhanced working memory capacity. They have proposed a number of possibilities, including increased phonological storage capacity, greater inhibitory function of the central executive, greater range of speech acts (the reasons why people speak), more powerful theory of mind (being able to understand what someone else is thinking), and/or some heretofore unexamined, domain-specific aspect of working memory. Archaeological support for this enhancement includes evidence for managed foraging systems that planned months and years in advance, age and gender divisions of economic behavior (in which adult men hunted big game, while women and juveniles foraged and hunted small game), depictive artistic traditions that included abstract concepts (e.g., half-animal, half-human figures), and the ability to plan and successfully conduct over-the-horizon colonizing voyages. The timing of this development hinges on the serendipity of archaeological discovery. A strict reading from multiple lines of evidence places enhanced working memory very late in human evolution, sometime after 50,000 years ago. A sanguine interpretation of more scattered evidence (e.g., beads made from shells) would push evidence for modern working memory back to 100,000 years ago.

Frederick L. Coolidge and Thomas Wynn

See also Attention, Evolutionary Perspectives; Visual Working Memory; Working Memory; Working Memory in Language Processing

Further Readings


**Working Memory in Language Processing**

Adult comprehenders differ in language-specific skills involved in processing words, sentences, and extended discourse. Skilled adult comprehenders also differ in general abilities, such as the ability to flexibly allocate attention, to suppress or inhibit irrelevant or distracting information, in overall processing speed, and in working memory capacity (WMC). Working memory is conceptualized as a cognitive organ in which information is kept in a readily accessible form and manipulated as needed. Some theories of language comprehension claim that working memory is the core ability that determines why some individuals process language more efficiently and effectively than others. This explains why working memory tasks correlate with measures of comprehension ability and overall verbal ability, whereas short-term memory tasks do not. This entry outlines three approaches to working memory that explain why differences in WMC can lead to differences in language comprehension ability. It also reviews evidence suggesting that WMC relates to comprehension ability in skilled adults only by virtue of its relations to other reader characteristics. Because of space limitations, it does not review WMC contributions to language production processes.

**How Is WMC Measured?**

To understand language, comprehenders must undertake multiple related processes, including lexical access, syntactic parsing, and contextual integration. As these processes are unfolding, readers must maintain their comprehension goals, information extracted from the text, world knowledge, and the partial products of interpretive processes. Working memory supports each of these functions. Working memory resembles short-term memory, which also keeps information temporarily active. Working memory differs from short-term memory because it entails both storage and processing. Working memory also includes executive processes that regulate and control task-relevant information.

Working memory is uncontroversially a limited capacity system, but different theories make different claims about the factors responsible for capacity
limitations. These factors may include limitations on the amount of activation, similarity-based interference, processing speed, encoding and retrieval problems, and the ability to inhibit irrelevant information.

Variation in WMC has been shown to correlate with performance on a range of cognitive tasks, including language comprehension. Complex language-processing tasks place the greatest demands on working memory and so offer the opportunity to observe differences in performance across individuals who differ in WMC. Complex tasks involve multiple component processes, however, which complicates the interpretation of any observed correlations between WMC and task performance.

WMC measures assess an individual's ability to keep information activated while undertaking a task that prevents rehearsal. Reading span is often used to measure WMC. In this task, participants read aloud a set of sentences, presented one at a time. They recall the final word of each sentence after reading the entire set. Reading span is the largest set size for which a reader recalls all the sentence-final words. Reading span correlates with the verbal Scholastic Aptitude Test (SAT) (about .5) and the ability to answer questions about a text (about .8). When reading span and text complexity are used to predict reading times, they interact. Differences between high- and low-capacity comprehenders increase as text complexity increases. However, reading span does not correlate with short-term memory tasks, such as digit span.

Other measures of working memory capacity include the following: (a) operation span—participants perform simple arithmetic problems while retaining a set of words; (b) alphabet span—participants repeat a list of words after arranging them in alphabetic order; and (c) minus span—participants repeat a list of numbers after subtracting two from each. These measures correlate highly with each other and with the reading-span measure.

The Relation Between WMC and Language Comprehension

Marcel Just and Pat Carpenter's theory of working memory and language comprehension requires a working memory system that maintains partial products of the comprehension process in an active state while additional input is analyzed. A general purpose working memory system supports both linguistic and nonlinguistic functions. Moreover, the storage and processing functions are fueled by activation, a commodity that maintains knowledge elements in memory and supports computation. Activation is shared among storage and processing functions such that activation-consuming processes limit the amount of activation available to support storage and vice versa. Individual differences in language comprehension depend on differences in capacity, the total amount of activation available to the system. Language places demands on limited working memory resources, and more complex language creates greater demands than less complex language. So, for example, syntactically complex sentences consume more memory resources than simple sentences. As a result, individuals with lower working memory capacities should have greater difficulty with more complex than simple sentences, whereas individuals with large working memory capacities should process complex sentences with about the same efficiency as they process less complex sentences. In some studies, groups of individuals with lower WMC show greater complexity effects than groups with higher WMC. Dual-task paradigms also provide evidence that lower capacity comprehenders are more affected by complexity than are higher capacity comprehenders.

Just and Carpenter's formulation has been challenged by two separate lines of inquiry. First, working memory capacity does not interact with syntactic complexity when appropriate statistical methods are used. Second, groups of patients with scores of zero on working memory tasks are able to parse and interpret sentences containing complex structures and long-distance dependencies. Results such as these have motivated David Caplan and Gloria Waters to propose two separate sources of working memory resources: One is involved in interpretive functions (e.g., lexical access, parsing, and the assignment of standard meaning), and the other is involved in post-interpretive functions (e.g., contextual integration, and inferencing). Aphasic patients could parse and interpret complex sentences even though they could not remember any words in the reading-span task, perhaps because one source of working memory resources was impaired while the other source was intact. Minimally, theories of WMC must explain how an individual whose reading-span score is zero can comprehend complex sentences.
Other approaches to working memory and language start by noting that scores on working memory tasks depend on (a) the content of the information being held in working memory and (b) the degree to which processes have been automatized. The reorganization of to-be-remembered information by chunking can increase the apparent capacity of working memory, so a single retrieval cue can recover vastly greater information than the standard seven plus or minus two “chunks.” Further, automating a process can greatly reduce the demands that processes impose on working memory. To determine the extent to which a task loads working memory, one has to know what information is being manipulated, how that information is organized, and how much practice the individual has had with that specific task. Language interpretation processes such as syntactic parsing are overlearned and automatized; thus, they may place minimal demands on working memory. This account places heavy emphasis on experience—the more often you have encountered a stimulus of a given type, the more automated the interpretive processes should be and the more efficient you should be in processing that stimulus.

A third approach to working memory and language appeals to processes that “clean up” the by-products of automated access and retrieval processes. These approaches argue that differences in working memory capacity are a by-product of the ability to efficiently activate relevant information and suppress or ignore activated but irrelevant information. Two individuals may have equivalent ability to activate and manipulate information, but one individual may overactivate associated information in response to a particular string of words or be unable to reduce or suppress information not directly relevant to the intended meaning. In that case, the individual will behave as though he or she has a small working memory, because the available capacity supports the activation of irrelevant information.

Methodological Advances in Studying WMC and Language

Research on the relation between WMC and language comprehension has been plagued by a set of methodological problems. These involve the psychometric qualities of the instruments used to measure working memory capacity, the nature of the statistical designs used to test the relation between working memory and language comprehension, and the fact that working memory capacity is correlated with a wide variety of other individual differences. Studies often rely on a single measure of working memory capacity. No single measure of capacity has exhibited very high reliability in studies assessing the psychometric properties of working memory tasks. In other words, tests used to assess WMC do not produce the same score for the same individual if the test is taken on two separate occasions. This problem can largely be avoided by using multiple measures of WMC.

The relation between WMC and language comprehension is often studied using quasi-experimental designs. WMC is measured and then people participate in an experiment in which some text variable is manipulated (e.g., word frequency, syntactic complexity). Quasi-experimental designs treat continuous variables (such as WMC) as categorical. This technique can artificially magnify differences between groups and mask variation in performance within each artificially established group. Most of the published quasi-experimental studies of individual differences on WMC and language comprehension also use an extreme-groups design. In such experiments, individuals are selected for analysis because they score very high or very low on some test, such as the reading-span test. Subjects who are closer to average are excluded. This kind of experiment does not allow one to draw conclusions about performance on the language comprehension task across the full range of WMC. Finally, the results of these studies can be difficult to interpret when WMC is the only measure of individual variation examined. Individuals who score low on measures of working memory span tend also to score low on tests of word-recognition ability, vocabulary, print exposure, reasoning ability, and domain knowledge. Is variation in language comprehension because of individual differences in WMC, or is it secondary to abilities that are correlated with capacity?

Some researchers have begun to examine this question using multiple regression and multilevel modeling techniques. For example, Bruce Britton and his colleagues used structural-equation modeling to examine the individual characteristics that affect learning from instructional texts. They assessed the influence of four individual difference factors on text learning: the ability to make inferences, metacognitive ability, working memory capacity, and domain
knowledge. Text learning was predicted by a reliable path such that metacognitive ability predicted inference-making ability; inference-making ability predicted domain knowledge, and domain knowledge predicted text learning. Also, metacognitive ability predicted inference ability and inference ability predicted WMC. The relation between WMC and text learning, however, was not significant. Thus, WMC strongly correlated with other individual differences but did not predict text learning when these other variables were entered into the model.

Other studies have also failed to find a relation between language comprehension and working memory capacity when correlated variables have been included in the analyses. Alexandra Gottardo, Linda Siegel, and Keith Stanovich examined the influence of WMC on comprehension in adults with reading disabilities. They found that WMC predicted reading comprehension when it was entered early in a regression equation but failed to predict unique variance when it was entered after other variables.

Debra Long and her colleagues found similar results in a study examining how individual difference variables interact with properties of texts to influence comprehension. They assessed participants’ performance on several information processing and language tasks. The individual difference tests included (a) word decoding speed (how quickly a person can pronounce a visually presented word), (b) word decoding accuracy (the number of word-reading errors), and (c) WMC. Sentence-reading times for each participant were analyzed as a function of three text characteristics: (a) number of function words (grammatical markers such as of and although); (b) number of new argument nouns—how many new concepts a sentence introduces; and (c) number of repeated argument nouns—the number of old concepts. Individual differences in reading times were influenced by different combinations of word-decoding ability, overall verbal ability, and print exposure. When all the individual difference variables were included in the analyses, working memory capacity did not moderate the effect of function words, repeated arguments, or new arguments. When WMC was entered into the model by itself, it did predict the relation between the number of function words and sentence-reading time.

Research examining the shared variance among working memory capacity and other individual-difference variables is still in its infancy. Initial studies suggest that the relation between WMC and language comprehension may be derivative. WMC and language comprehension share variance because WMC is correlated with other variables that have a causal relation with comprehension. More studies are needed before we can definitively say whether WMC plays a direct role in explaining variation in language comprehension ability.

Matthew J. Traxler and Debra L. Long

See also Attention, Resource Models; Automaticity; Discourse Processing, Models of; Working Memory

Further Readings


Index

A
A not B error, 2:715
Aamodt, Agnar, 1:129
Aarts, Henk, 1:192
Ability, 1:84, 2:531
Abnormal Spindle-like, Microcephaly-associated protein (ASPM), 1:371–372
Aboutness, 1:140, 187, 198, 199, 201, 240, 426.

See also Intention and intentionality
Abstract letter representations, 2:785
Abstract words, 1:115
Acamprosate, 1:10–11
Accents, 2:621
Acceptance and commitment therapy, 1:107
Access consciousness, 1:1–3, 321, 2:561
Accessibility effects, 2:618
Accessibility theories, of spacing effect, 2:711
Accuracy
motivation to achieve, 2:528
promotion of, 2:694
reaction time and, 2:632–633
speed trade-off with, 1:220, 2:694, 763
Acquaintance, 2:452–453
Acronyms, as mnemonic device, 2:523
Action
attention and, 1:62–65
basic actions, 1:4–5
broad vs. narrow senses, 1:330
causal theory of, 2:597–599
collective action, 1:158–160
defined, 1:4, 330
desire and, 1:244, 246
events and, 1:4–5, 2:597
explanation of, 1:330–333
freedom of, 1:360–363
mental action, 2:497–498
mirror neurons, 2:520–521
perception and, 1:161, 2:595–596
phenomenology of, 2:595–596
philosophy of, 2:596–600
prospective, 2:533–534
reasons for, 2:599–600
teleological vs. causal explanations, 2:735–737

Action and bodily movement, 1:4–5, 46–50
Action disorganization syndrome, 1:8
Action schemas, 1:7
Action selection, 2:556–558
Action simulation, 1:162
Action slips, 1:5–8
Action theory, 2:443
Action understanding, 2:443
Adams, Fred, 1:337
Adams, John, 1:326
Adams, Susan, 1:210
Adolescence
identity, development of, 2:586
relationships, development of, 2:649–650
schizophrenia, 2:664–665
self, development of, 2:670–671
Adolphs, Ralph, 1:317
Adoption studies, 1:10
Adorno, Theodor, 2:609–610
Adulthood, word learning in, 2:782
Advertising, 2:724, 753, 754
Affect, 1:84–85, 298–300, 305
Affect-backed rules, 1:302
Affective circumplex, 1:298–299
Affective consciousness, of animals, 1:190–193
Affective forecasting, 1:11–13
Afferent representations, 1:161
Affordances, 1:197
Aftereffects, 1:344, 2:478–480
Aftertastes, 2:733
Agape (love), 2:473
Agency, 2:772–773
Agent causation, 1:363, 2:598
Aggression, 1:324
Aging, memory, and information processing speed, 1:13–18
  measures, 1:13–14
  memory stages, 1:14–17
  reaction time, 1:17–18
Agnosia, 1:346
Agoraphobia, 1:40
Agrammatism, 1:44
Agraphia, 1:44
Agreement, grammatical. See Syntactic production, agreement in
Ainsworth, Mary, 2:647
Aizawa, Ken, 1:337
Ajzen, Icek, 1:79–81
Akinetopsia, 1:123
Akrasia (weakness of will), 2:596
Alcohol abuse, 1:9
Alcohol dependence, 1:9
Alexander, Joshua, 1:330
Alicke, Mark, 1:329
Alignment-based models of similarity, 2:690
Allais, Maurice, 1:19, 222, 2:565
Allais paradox, 1:19–21, 222, 2:565
Allen, John, 1:289
Allocentric mental representations, 2:771
Allport, Alan, 1:61, 260
Allport, Gordon, 1:79
Alpha rhythm, 1:274
Alphabet span, 2:794
Altemeyer, Robert, 2:610
Altruism, 1:248, 301
Alvarez, Maria, 1:5
Alzheimer's disease
  action slips, 1:8
  confabulation, 1:184
  exercise and the brain, 1:327
  self-consciousness, 2:674
Ambiguity, 1:216, 237
American Psychological Association, Task Force on Statistical Inference, 2:507
Ames room, 1:232–233
Anamnesis, 1:22–25
  anterograde, 1:23–24, 95, 400
  automatism, 1:95
  brain, 2:484–487
  etiologies, 1:22–23
  infantile/childhood, 1:326
  Korsakoff's amnesia, 1:183
  neuroanatomical correlates, 1:22–23
  retroactive interference, 2:481–482
  retrograde, 1:24–25, 95, 2:485
  Amodal theories, 1:174–175
  Amount-of-processing hypothesis, 2:661
  Amnesia, 1:346
  Aggression, 1:324
  Anomalous monism, 1:32–34, 2:500
  Anaphors, 1:408
  Anchoring, 1:28–29
  Anderson's problem, 1:175
  Antisocial personality disorder, 2:555
  Anxiety, 1:40
Anxiety disorders, 1:39–43
comorbidity, 1:42
diagnosis, 1:40
etiology, 1:42
overview, 1:39–40
prevalence, 1:39–40
treatment, 1:42–43
types, 1:40–42
Anxious-ambivalent attachment style, 2:475, 647–648

Aphasia, 1:43–46
apraxia and, 1:47
Broca's aphasia, 1:44
clinical syndromes, 1:43–44
conduction aphasia, 1:44, 181–183
global aphasia, 1:44, 279
jargon aphasia, 1:36
neurological correlates, 1:45–46
overview, 1:43
primary progressive aphasia, 1:45–46
psycholinguistic approach, 1:44–45
pure motor speech disorders, 1:44
pure word deafness, 1:44
tanscortical motor aphasia, 1:44
tanscortical sensory aphasia, 1:44
Wernicke's aphasia, 1:44
word learning, 2:782

Appearing, theory of, 2:737–739
Applied behavior analysis, 1:93
Appraisal theories, 1:166
Appraisals, 1:429, 431
Approach behaviors, 1:288–289

Apraxia, 1:46–50
Aquinas, Thomas, 2:597
Area Spt, 1:183
Arginine vasopressin, 1:248
Argument from illusion, 2:452
Argument mapping, 1:50–52
Arico, Adam, 1:328
Aristotle, 1:135, 147–149, 301, 324, 2:499, 596–597, 608, 660, 674, 739
Aristoxenus, 2:549
Armstrong, David, 2:676
Armstrong, J. Scott, 2:777
Aronson, Elliot, 1:154
Arrow's Impossibility Result, 1:224
Articulation, place and manner of, 2:717
Articulatory suppression, 2:642
Artificial intelligence (AI), 1:129–130, 280
Artificial neural networks, 2:645–646
“As if” models, 2:565
Aschersleben, Gisa, 1:162
Aserinsky, Eugene, 2:695
Asomatopsia, 1:101
Asperger's disorder, 1:90
Association for Behavior Analysis International, 1:110
Associative asymmetry, 2:494
Associative learning, 1:78, 170
Associative memory, 2:487–488
Associative networks, in social cognition, 2:704–705
Astronomy, 2:630
Asymmetric dependence, 1:140–141
Asymmetry effect, 2:493–494
Atomism about concepts, 1:52–53, 172–173. See also Classical theory of concepts; Concepts, philosophical issues
Attachment styles, 2:442, 475, 647–650
Attention
autism, 1:90–91
automaticity, 1:96
change blindness, 1:146
control of, 1:58–59
deployment, 1:312
disengaging and reorienting, 1:59
divided, 1:260–263
endogenous, 1:68
endowment effect, 1:322
exogenous, 1:68
inattentional blindness, 1:403–406
instruction, 2:530–531
internal vs. external, 2:531–532
memory, 1:401–402
neurophysiology of, 2:563–564
perceptual consciousness, 2:577–580
performance limits, 1:60–62, 74–75
selective attention, 1:53–56, 75
tsites of, 1:57–58
Stroop effect, 2:723
visual, 2:562–564
visual search, 2:763, 765
working memory, 1:62–65
Attention, evolutionary perspectives, 1:53–56
Attention, neuroimaging studies of, 1:56–59, 2:723
Attention, resource models, 1:60–62, 75
Attention and action, 1:62–65, 2:543–546. See also Divided attention and memory
Attention and consciousness, 1:66–68
Attention and emotion, 1:68–72, 312
Attention and emotions, computational perspectives, 1:72–73
Attentional blink hyperactivity disorder (ADHD), 1:125
Attentional capture, 1:59, 64, 68, 406, 410, 411
Attentional focus, 2:531–532, 693, 788
Attentional load, 2:578
Attentional orienting, 1:411
Attentional sets, 1:405
Attitude change, 1:76–78, 153. See also Persuasion
Attitude strength, 1:80
Attitudes
  internal structure, 2:708
  persuasion, 2:590–594
  pro-attitudes, 1:331–332
  social cognition field, 2:706
  stability of, 1:76–77
Attitudes and behavior, 1:76–78, 79–81. See also Personality: individual versus situation debate
Attraction, 1:82–83, 345
Attraction errors, 2:461–462
Attribute inheritance model, 1:180
Attribution theory, 1:83–85, 2:587. See also Causal theories of intentionality
Attribution theory, 2:706
Audience design hypothesis, 2:589
Audition, neural basis, 1:85–88
  brain stem processing, 1:86–87
  nonspatial perception, 1:87–88
  peripheral processing, 1:85–86
  spatial perception, 1:87
  thalamic processing, 1:87
Auditory masking, 1:88–90. See also Masking
  Auditory nerve, 1:384
  Auditory scene analysis, 1:386
  Authoritarianism, 2:609–610
Authority
  persuasion, 2:593
  self-knowledge, 2:676–677
Authorship processing, 2:773
Autism, 1:90–93
  assessment, 1:92–93
  characteristics, 1:90–91
  comorbidity, 1:90
  developmental course, 1:91
  diagnosis, 1:90
  epidemiology, 1:91
  pathophysiology, 1:91–92
  treatment, 1:93
Autism spectrum disorders (ASDs), 1:90, 2:459
Autobiographical memory, 1:24, 2:678, 679, 759
Autobiographical self-consciousness, 2:674
Automated essay grading, 2:682
Automatic behavior, 1:93–95
Automatic processes, 2:707
Automatic speech recognition, 2:477–478
Automaticity, 1:95–98
  categorization, 1:133–134
  conceptual issues, 1:97–98
  criteria, 1:96–97
  methodological issues, 1:98
  reading, 2:722
  working memory, 2:795
Autonoetic consciousness, 1:197
Autonomy, 2:526
Autopoietic theory, 1:197
Autoscopic phenomena, 1:99–102
  hallucinations, 1:99–100
  heautoscopy, 1:100
  out-of-body experiences, 1:100–101
  presence, 1:101
  rare types, 1:101
Availability heuristic, 1:102–103
Avoidant attachment style, 2:475, 647
Awareness
  anesthesia, 1:30–31
  anosognosia, 1:34–37
  automaticity, 1:97
  defined, 2:515
  subliminal perception, 2:723–725
Ayer, A. J., 1:398
Baars, Bernard, 1:3
Backchannels, 1:204
Background knowledge, 1:180
Backmasking, 2:753
Backward masking, 1:123, 2:753, 760, 761. See also Masking
Baddeley, Alan, 1:14–15, 261, 422, 2:641, 766, 786, 789–792
Bahrick, H. P., 1:16
Baird, John, 2:582
Balanced anesthesia, 1:30
Ball, Linden, 2:740
Ballistic processes, 1:97
Banks, Martin, 2:574
Barlow, Horace, 2:721
Barrouillet, Pierre, 2:643, 789
Barsalou, Lawrence, 1:163, 174–175
Barto, Andrew, 1:218
Bas relief depth, 1:234
Basal ganglia, 1:246, 393, 2:557, 558
Base-rate neglect, 2:657–658
Basic actions, 1:4–5
Basic level, of categorization, 1:136
Basilar membrane, 1:85, 383
Batchelder, William, 2:539
Batson, Daniel, 1:301
Bauer, Russell, 1:314
Baum, William, 1:110
Bauman, Margaret, 1:91
Baumeister, Roy, 2:502
Bayesian models, 1:139
Bayesian rationality, 2:742
Bayesianism, 1:264
Beall, Andrew, 2:576
Bearing, 2:576
Beat gestures, 1:373
Beck, Aaron, 1:42, 106–107
Beecher, Henry, 2:604
Behavior
attitudes and, 1:79–81
congruent and incongruent, 2:704–705
teleological vs. causal explanations, 2:735–737
Behavioral decision theory, 1:321–323
Behavioral economics, 2:565
Behavioral sequencing, 2:557–558
Behavioral therapy, 1:105–108
acceptance and commitment therapy, 1:107
cognitive behavioral therapy, 1:106–107
mindfulness based cognitive therapy, 1:107
origins, 1:105–106
Behaviorism, 1:108–111
consciousness and embodiment, 1:195
decision making, 1:218
language development, 2:455
mind-body problem, 2:517
reductive physicalism, 2:638
reinforcement learning, 2:643–646
thinking, 2:739
Beilock, Sian, 2:747
Bekkering, Harold, 1:162
Bekoff, Marc, 1:191
Belief
concepts and language, 1:177
degrees of, 1:264–265
explanation of action, 1:331, 2:599–600
instrumentalism, 2:634–635
justified, 1:229–230
memory and, 2:490–491
persuasion, 2:590–594
Belief and judgment, 1:111–114. See also Belief; Judgment
analyzing belief, 1:113–114
attitude change, 1:76–78
belief vs. judgment, 1:111–112
debiasing, 1:207–209
epistemology, 1:112–113
propositions, 1:111
Belief bias, 2:742
Belief disconfirmation paradigm, 1:154
Bell, Charles, 1:352
Bell Laboratories, 1:85
Bellman, Richard (Dick), 1:213, 218
Bellugi, Ursula, 1:416
Bem, Daryl, 1:110, 437
Bemporad, Brenda, 1:289
Bentall, Richard, 1:231
Bergson, Henri, 2:513
Berkeley, George, 1:397–398, 2:548–549
Berko, Jean, 2:458
Bermúdez, José, 1:360
Bernard, Claude, 1:109
Bernoulli, Daniel, 1:20, 221
Berridge, Kent, 1:10, 245, 246
Berwick, Robert, 1:414
Bialystok, Ellen, 1:116
Bias
belief bias, 2:742
children’s knowledge, 2:451
debiasing, 1:207–209
decision making, 1:212
dissent as antidote to, 1:257
hard-wired, 1:209
learning, 1:139
love, 2:474–475
matching bias, 2:740–741
researcher, 2:610
word learning, 2:781
Biased accessibility, 1:29
Biased competition model, 1:55, 2:563
Bias-variance dilemma, 1:138–139
Bickerton, Derek, 2:547
Bilgrami, Akeel, 2:677
Bilingual language processing, 1:114–115
Bilingualism, 1:391–392, 2:458–459, 782
Bilingualism, cognitive benefits of, 1:115–118
Bin model of information access, 2:706
Binet, Alfred, 1:422
Binocular disparity, 2:719
Binocular rivalry, 2:578–579
Biology
children’s theories of, 2:451
love, 2:474–475
reinforcement, 2:645
Bipolar disorder, 1:125
Birdsong, 2:547
Birdwhistell, Ray, 1:352
Birren, Jim, 1:18
Bispectral analysis (BIS), 1:31
Bispectral Index, 2:754
Bistable perception, 2:561
Bjork, Robert, 1:242, 2:711
Black English Vernacular (Ebonics), 1:118–120
Blackburn, Simon, 1:240
Blair, James, 1:303–304
Blake, Randolph, 2:721
Blalock, Hubert M., 1:434–435
Bleuler, Paul Eugen, 2:663
Blindsight, 1:120–123
access consciousness and, 1:2, 3
attention and consciousness, 1:68
controversies, 1:121

cortical blindness, 1:120–121
discovery of, 1:121
limits, 1:122
neurological explanation, 1:122–123
normally sighted persons, 1:123
true blindsight, 1:121–122
unconscious perception, 2:754
Bliss, Timothy, 2:645
Block, Ned, 1:1–3, 172
Bloomfield, Leonard, 1:119
Bobrow, Daniel, 1:60
Bock, Kathryn, 2:461, 462

Body. See also Mind-body problem
actions and, 1:4–5
apraxia, 1:46–50
attraction, 1:82
autoscopic phenomena, 1:99–102
confabulation, 1:185
consciousness, 1:194–198
embarrassment, 1:282
emotional expression, 1:292
intentionality and sensation, 1:427–428
knowledge-how, 2:447–448
language development, 2:455–456
lie detection, 2:470–472
spatial mental representations, 2:770–771
Boghossian, Paul, 1:230
Bond, Alan, 1:54
Bonhoeffer, Karl, 1:183
Bootstrapping models of prediction, 2:614
Borderline personality disorder, 1:124–125, 2:555
Boring, Edwin, 2:581
Borod, Joan, 1:288, 314
Bottlenecks. See Processing bottlenecks
Bounded rationality, 1:212
Bowers, Dawn, 1:314
Bowlby, John, 1:307, 2:647
Bowman, Howard, 1:73
Brady, Michael, 2:629
Brain. See also Left hemisphere of brain; Neuroimaging;
Neurology; Right hemisphere of brain
action selection, 2:557
aging, 1:18–19
analogical reasoning, 1:26
anosognosia, 1:34–37
attention, 1:54, 56–59, 75
autism, 1:91–92
automatisms, 1:94
blindsight, 1:120, 122–123
category learning, 1:130–132
cognition, 1:279
conceptualization, 1:175
consciousness, 2:513–515, 560–562
decision making, 1:215–217, 248
discovery of, 1:121
limits, 1:122
neurological explanation, 1:122–123
normally sighted persons, 1:123
two blindsight, 1:121–122
unconscious perception, 2:754
Bliss, Timothy, 2:645
Block, Ned, 1:1–3, 172
Bloomfield, Leonard, 1:119
Bobrow, Daniel, 1:60
Bock, Kathryn, 2:461, 462
Bregman, Albert, 1:386
Breiter, Hans, 2:366
Breitmeyer, Bruno, 2:762
Bremmer, Frank, 2:575
Brennan, Susan, 2:621
Brentano, Franz, 1:201, 426, 2:674
Bridging inferences, 1:408–409
Brinley, Joseph, 1:17
British Empiricism, 2:739
Britton, Bruce, 2:795
Broad, C. D., 1:286
Broadbent, Donald, 1:60, 64
Broca’s aphasia, 1:44, 127, 2:456, 462–463
Broca’s area, 1:224, 277, 279, 2:456
Brodmann’s areas, 2:475
Broca’s aphasia, 1:224
Brody, Nathan, 2:752
Brogaard, Berit, 607
Brody, Vicki, 1:350
Brunswik, Egon, 2:581
Buccofacial apraxia (BFA), 1:48
Buck, Ross, 1:315
Buddhism, 2:629
Budescu, David, 2:778
Bullock, August, 2:753
Buprenorphine, 1:10
Buproprion, 1:11
Burge, Tyler, 1:37, 38, 39, 152, 2:571
Burgess, Neil, 2:686
Burr, David, 2:575
Buss, David, 1:323, 2:475
Butler, Joseph, 2:639
Cacioppo, John, 2:591
Callosal apraxia, 1:48
Caloric, 1:278
Caltagirone, Carlo, 1:314
Camblin, Christine, 1:58
Camerer, Colin, 2:566–567
Cameron, Deborah, 1:370
Camouflage, 1:234
Campbell, John, 1:255, 2:453
Canaleto, 1:233
Cannon, Walter, 1:109
Capgras, Jean Marie Joseph, 1:127
Capgras delusion, 1:127–128, 185, 231, 364–365
Caplan, David, 2:794
Capture errors, 1:6, 7
Capture of attention. See Attentional capture
Caramazza, Alfonso, 1:174, 175
Cardozo, Benjamin, 2:469
Caregiving system, 2:475
Carey, Susan, 1:172
Carlson, V. R., 2:582
Carpenter, Pat, 2:794
Carrier, Mark, 1:262
Carroll, Lewis, 1:343, 408
Carston, Robyn, 2:511
Casanova, Manuel, 1:91–92
Case-based reasoning, 2:468, 469
Case-based reasoning, computational perspectives, 1:128–130
Case-marking languages, 2:618
Categorical perception, 2:718
Categories. See also Human classification learning
ad hoc, 1:135–136
basic level, 1:136
coherence of, 1:135–136
defined, 1:134
dimensionality, 1:139–140
functions, 1:134
goal-based view, 1:136–137
human classification learning, 1:392–394
intuitive vs. deliberative reasoning, 2:746–747
organization of knowledge about, 1:134–136
type-based view, 1:135–136
types, 1:137–138
Categorization, neural basis, 1:130–134. See also
Memory, neural basis
automatic categorization, 1:133–134
category-learning systems, 1:130
information-integration vs. rule-based, 1:132
other brain areas, 1:133
perceptual representation memory system, 1:133
prefrontal cortex, 1:130, 132
striatum, 1:132–133
See also Concepts, philosophical issues
Category accessibility, 2:706
Category learning, computational perspectives, 1:138–140
Cattell, James McKeen, 2:722
Causal completeness of the physical, 2:499–500
Causal decision theory, 1:223
Causal graphs, 2:523–524
Causal judgments, 2:524
Causal learning, 2:523–525
Causal Markov assumption, 2:524
Causal theories of intentionality, 1:140–143. See also
Attribution theory; Intentionality of bodily sensation; Intentionality of emotion
Causal theories of memory, 1:144–145
Causal theory of action, 2:597–599
Causal theory of appearing, 2:738
Causality
- anomalous monism, 1:32–33
- explanation of action, 1:332, 2:597–599
- mental causation, 2:499–501
- mind-body problem, 2:516–517
- teleology and, 2:735–737
- CBS (television network), 1:213
- Central executive processes, 1:15, 2:791–792
- Central processing, 1:64, 2:544, 546, 591, 623.

See also Serial processing
- Central storage theory, 2:790
- Cerebellum, 2:488
- Cerebral cortex
  - evolution, 1:54
  - hearing, 1:87
  - memory, 2:485
  - selective attention, 1:54
- Cerebral vascular accidents, 1:23
- Cerella, John, 1:17
- Certainty effect, 1:21
- Chabris, Christopher, 2:577
- Chaiken, Shelly, 2:591
- Chaining, 2:684, 685
- Chalmers, David, 1:3, 286, 335–337, 2:453
- Chameleon effect, 2:521
- Change blindness, 1:67, 145–147, 405–406, 2:577–578
- Character, and moral development, 2:525–527
- Character and personality, philosophical perspectives,
  1:147–150. See also Personality: individual versus situation debate
- Characteristic frequency (hearing), 1:384
- Chater, Nick, 2:690, 742
- Cheesman, Jim, 2:751
- Chen, Ji-Qie, 2:543
- Cheng, Patricia, 2:229
- Chicken (game), 1:160
- Childhood disintegrative disorder, 1:90
- Children
  - analogical reasoning, 1:26–28
  - attachment, 2:647
  - conceptualization by, 1:168–171
  - confabulation, 1:184
  - dyslexia, 1:267–271
  - emotional expression, 1:291
  - face perception, 1:344–345
  - folk psychology, 1:359
  - identity development, 2:586
  - inner speech, 1:418–419
  - jealousy, 2:441–442
  - knowledge acquisition, 2:449–451
  - language development, 1:413–417, 2:455–459, 462
  - memory, 1:325–326
  - moral development, 2:525–527
  - representation development, 2:654–656
- schizophrenia, 2:663–664
- scientific theories, 2:450–451
- self, development of, 2:669–671
- spatial cognition, 2:713–716
- word learning, 2:779–781
- Chill response, 2:550
- Chinese language, 1:180–181
- Chinese room argument, 2:653–654
- Chloroform, 1:30
- Choice reaction time tasks, 2:631–632
- Chomsky, Noam, 1:151, 2:455
- Choshen-Hillel, Shoham, 2:778
- Christensen, David, 1:265
- Christianity, 2:629
- Chunks, 2:789, 795
- Churchland, Paul, 1:358, 2:634–636
- Cialdini, Robert, 2:590, 592
- Cicero, Marcus Tullius, 1:326, 2:522
- Claparede, Édouard, 1:400
- Clark, Andy, 1:335–337
- Clark, Herbert H., 1:203
- Class conflict, 1:434
- Classical conditioning, 1:105, 2:605–606, 643–644, 743
- Classical fear conditioning, 1:293–294
- Classical foundationalism, 2:489–490

See also Atomism about concepts; Concepts, philosophical issues
- Classification. See Human classification learning
- Clemen, Robert, 2:777
- Clinical high risk research, 2:665
- Clinical prediction, 2:613–615
- Closure, motivation to achieve, 2:528
- Coarticulation, 2:717
- Cochlea, 1:85
- Cocktail party effect, 1:66
- Cocteau, Jean, 1:101–102
- Coenen, Volker, 1:307
- Coextensive concepts, 1:53
- Cognates, 1:115
- Cognition
  - anti-individualism, 1:37–39
  - bilingualism, 1:115–118
  - brain, 1:279
  - distributed, 1:258–260
  - embodied, 1:163, 351
  - evolution, 2:792–793
  - exercise, 1:326–328
  - individualism, 1:37
  - language, 1:278–279
  - multinomial modeling, 2:538–541
  - psychopathologies, 1:304–308
  - psychophysiological reactivity, 1:296–297
  - spatial, 2:713–716
<table>
<thead>
<tr>
<th>term</th>
<th>page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Companionate love</td>
<td>2:474</td>
</tr>
<tr>
<td>Common ground</td>
<td>2:591</td>
</tr>
<tr>
<td>Cognitive architectures, 1:165–166</td>
<td>1:165–166</td>
</tr>
<tr>
<td>Cognitive behavioral theory, 1:10</td>
<td>1:10</td>
</tr>
<tr>
<td>Cognitive behavioral therapy</td>
<td>2:42–43</td>
</tr>
<tr>
<td>anxiety disorders, 1:42–43</td>
<td>1:125</td>
</tr>
<tr>
<td>borderline personality disorder, 1:10</td>
<td>1:106–107</td>
</tr>
<tr>
<td>Cognitive control, 2:504</td>
<td>2:504</td>
</tr>
<tr>
<td>Cognitive dissonance, 1:78, 153–156</td>
<td>1:153–156</td>
</tr>
<tr>
<td>Cognitive ecology, 1:259</td>
<td>1:259</td>
</tr>
<tr>
<td>Cognitive integration, 1:335</td>
<td>1:335</td>
</tr>
<tr>
<td>Cognitive maps, 2:771</td>
<td>2:771</td>
</tr>
<tr>
<td>Cognitive neuroscience, 2:566, 668</td>
<td>2:566, 668</td>
</tr>
<tr>
<td>Cognitive psychology</td>
<td>2:666</td>
</tr>
<tr>
<td>debiasing, 1:208</td>
<td>1:208</td>
</tr>
<tr>
<td>deductive reasoning, 1:227–229</td>
<td>1:227–229</td>
</tr>
<tr>
<td>naive realism, 2:553</td>
<td>2:553</td>
</tr>
<tr>
<td>thinking, 2:739–740</td>
<td>2:739–740</td>
</tr>
<tr>
<td>unconscious, the, 1:200</td>
<td>1:200</td>
</tr>
<tr>
<td>Cognitive reappraisal, 1:254, 313</td>
<td>1:254, 313</td>
</tr>
<tr>
<td>Cognitive reserve, 1:116</td>
<td>1:116</td>
</tr>
<tr>
<td>Cognitive schemas, 1:106–107</td>
<td>1:106–107</td>
</tr>
<tr>
<td>Cognitive theories</td>
<td>2:448</td>
</tr>
<tr>
<td>intentionalty, 1:430–431</td>
<td>1:430–431</td>
</tr>
<tr>
<td>Cognitive-behavioral therapy, 2:666</td>
<td>2:666</td>
</tr>
<tr>
<td>Cohen, Geoffrey, 2:610</td>
<td>2:610</td>
</tr>
<tr>
<td>Cohen, J. D., 1:63</td>
<td>1:63</td>
</tr>
<tr>
<td>Coherence theory of truth, 1:198</td>
<td>1:198</td>
</tr>
<tr>
<td>Cohort model of auditory word recognition, 1:156–158, 2:784</td>
<td>1:156–158, 2:784</td>
</tr>
<tr>
<td>See also Word recognition, auditory</td>
<td>2:784</td>
</tr>
<tr>
<td>Coleman, Sam, 2:448</td>
<td>2:448</td>
</tr>
<tr>
<td>See also Joint or collective intention</td>
<td>1:158–160</td>
</tr>
<tr>
<td>Collective action dilemmas, 1:159–160</td>
<td>1:159–160</td>
</tr>
<tr>
<td>Color constancy, 2:580</td>
<td>2:580</td>
</tr>
<tr>
<td>Coltheart, Max, 1:231</td>
<td>1:231</td>
</tr>
<tr>
<td>Commitment</td>
<td>2:473, 475</td>
</tr>
<tr>
<td>interpersonal relationships, 2:473, 475</td>
<td>2:473, 475</td>
</tr>
<tr>
<td>joint intentionality, 2:444–445</td>
<td>2:444–445</td>
</tr>
<tr>
<td>persuasion, 2:594</td>
<td>2:594</td>
</tr>
<tr>
<td>Common environmental variance, 1:389–390</td>
<td>1:389–390</td>
</tr>
<tr>
<td>Common ground, 2:616–617</td>
<td>2:616–617</td>
</tr>
<tr>
<td>Companionate love, 2:474</td>
<td>2:474</td>
</tr>
<tr>
<td>See also Incompatibilism</td>
<td>1:329</td>
</tr>
<tr>
<td>Compatibility effects, 1:161–162</td>
<td>1:161–162</td>
</tr>
<tr>
<td>Compensation for coarticulation, 2:717</td>
<td>2:717</td>
</tr>
<tr>
<td>Compensatory decision making, 1:216–217</td>
<td>1:216–217</td>
</tr>
<tr>
<td>Competitive queuing, 2:684–685</td>
<td>2:684–685</td>
</tr>
<tr>
<td>Complete love, 2:474</td>
<td>2:474</td>
</tr>
<tr>
<td>Compliance, 2:591</td>
<td>2:591</td>
</tr>
<tr>
<td>Componental model, 1:314</td>
<td>1:314</td>
</tr>
<tr>
<td>Compound words, processing of, 1:163–165</td>
<td>1:163–165</td>
</tr>
<tr>
<td>Computational auditory scene analysis, 1:386</td>
<td>1:386</td>
</tr>
<tr>
<td>Computational models. See also Modeling</td>
<td>1:386</td>
</tr>
<tr>
<td>attention and emotion, 1:72–73</td>
<td>1:72–73</td>
</tr>
<tr>
<td>category learning, 1:138–140</td>
<td>1:138–140</td>
</tr>
<tr>
<td>causal learning, 2:523–525</td>
<td>2:523–525</td>
</tr>
<tr>
<td>compound word processing, 1:163–165</td>
<td>1:163–165</td>
</tr>
<tr>
<td>consciousness and embodiment, 1:195, 197</td>
<td>1:195, 197</td>
</tr>
<tr>
<td>decision making, 1:218–220</td>
<td>1:218–220</td>
</tr>
<tr>
<td>discourse processing, 1:250</td>
<td>1:250</td>
</tr>
<tr>
<td>distributed cognition, 1:259–260</td>
<td>1:259–260</td>
</tr>
<tr>
<td>eye movements, 1:339</td>
<td>1:339</td>
</tr>
<tr>
<td>facial expressions, 1:349–351</td>
<td>1:349–351</td>
</tr>
<tr>
<td>multimodal modeling, 2:538–541</td>
<td>2:538–541</td>
</tr>
<tr>
<td>reinforcement, 2:645–646</td>
<td>2:645–646</td>
</tr>
<tr>
<td>scientific reasoning, 2:668</td>
<td>2:668</td>
</tr>
<tr>
<td>semantic memory, 2:681–683</td>
<td>2:681–683</td>
</tr>
<tr>
<td>sequential memory, 2:683–685</td>
<td>2:683–685</td>
</tr>
<tr>
<td>visual search, 2:563–564</td>
<td>2:563–564</td>
</tr>
<tr>
<td>Computational models of emotion, 1:72–73, 165–166</td>
<td>1:72–73, 165–166</td>
</tr>
<tr>
<td>Computer technology</td>
<td>1:50–52</td>
</tr>
<tr>
<td>argument mapping, 1:50–52</td>
<td>1:50–52</td>
</tr>
<tr>
<td>consciousness and embodiment, 1:195</td>
<td>1:195</td>
</tr>
<tr>
<td>conversation, 1:204</td>
<td>1:204</td>
</tr>
<tr>
<td>debiasing, 1:208</td>
<td>1:208</td>
</tr>
<tr>
<td>decision making, 1:211–214</td>
<td>1:211–214</td>
</tr>
<tr>
<td>face recognition, 1:347–349</td>
<td>1:347–349</td>
</tr>
<tr>
<td>multimodal conversational systems, 2:535–538</td>
<td>2:535–538</td>
</tr>
<tr>
<td>natural language generation, 2:558–560</td>
<td>2:558–560</td>
</tr>
<tr>
<td>rationality, 1:187</td>
<td>1:187</td>
</tr>
<tr>
<td>Computer-human interaction. See Human-computer interaction</td>
<td>1:211</td>
</tr>
<tr>
<td>Computerized decision aids, 1:211</td>
<td>1:211</td>
</tr>
<tr>
<td>Concealed information test, 2:471–472</td>
<td>2:471–472</td>
</tr>
<tr>
<td>Concept atomism, 1:52–53</td>
<td>1:52–53</td>
</tr>
<tr>
<td>Concept eliminativism, 1:174</td>
<td>1:174</td>
</tr>
<tr>
<td>Concept learning, 1:53</td>
<td>1:53</td>
</tr>
<tr>
<td>Concept mapping, 1:50</td>
<td>1:50</td>
</tr>
<tr>
<td>Concept pluralism, 1:174</td>
<td>1:174</td>
</tr>
<tr>
<td>Concepts</td>
<td>1:170</td>
</tr>
<tr>
<td>classical theory of, 1:150–152</td>
<td>1:150–152</td>
</tr>
<tr>
<td>coextensive, 1:53</td>
<td>1:53</td>
</tr>
<tr>
<td>demonstrative, 2:603</td>
<td>2:603</td>
</tr>
<tr>
<td>empiricst approach, 1:174–175</td>
<td>1:174–175</td>
</tr>
<tr>
<td>indexical, 2:603</td>
<td>2:603</td>
</tr>
<tr>
<td>innate, 1:170</td>
<td>1:170</td>
</tr>
</tbody>
</table>
Confabulation, 1:170
phenomenal, 1:286–287, 2:603
philosophical vs. psychological approaches, 1:171–172
semantic knowledge, 2:680–682
significance of, 1:168–169

Concepts, comparative perspectives, 1:166–168
Concepts, development of, 1:168–171.

See also Knowledge acquisition in development; Representations, development of
Concepts, philosophical issues, 1:171–175.

See also Categorization, psychological perspectives
classical theory of concepts, 1:150–152
conceptual atomism, 1:52–53
nonhumans, 1:166–168
Concepts and language, 1:169, 176–179
Conceptual combination, 1:179–181
Conceptual learning, 1:250–251
Conceptual representation, 2:654–656
Concrete words, 1:115
Conditional-response probability as a function of lag (lag-CRP), 2:493–494
Conditioned reinforcers, 2:645
Conditioning, 2:644
Conditioning theory, 2:605–606
Conduction aphasia, 1:44, 181–183
Confabulation, 1:183–185
Conflict. See Intergroup conflict; Intergroup conflict, models of
Congruent behavior, 2:704–705
Conjunction fallacy, 2:658, 747
Conjunction search, 2:562
Connectionism, 2:653
Conover, Kent, 2:566
Conrad, Reuben, 1:15
Conscious free will, 2:611–612

Conscious thinking, 1:186–189

Consciousness. See also Self-consciousness
affective, 1:190–193
anesthesia, 1:30–31
animals, 1:190–193
attention and, 1:66–68
blindsight, 1:122
brain and, 2:513–515
cultural construction of, 1:2–3
defined, 2:515
diagnosis, 1:320–321
introspection, 1:436–439
levels of, 1:191
microgenesis of, 2:513–515
naive realism, 2:553–554
neural correlates, 2:560–562
perceptual, 2:577–580
phenomenal, 1:1–3, 2:561–562, 602
sleep, 2:699
voluntary action, 2:773
wakeful, 2:498

Consciousness, comparative perspectives, 1:189–193

Consciousness and embodiment, 1:194–198. See also Personal identity
embodiment approach, 1:196–197
religious background, 1:194–195
scientific approaches, 1:195

Consciousness and the unconscious, 1:199–201
neural correlates, 2:560–562

Consciousness studies, 1:439
Conservatism, 2:610
Consistency
emotion, 2:629–630
persuasion, 2:594

Constancy, in emotion, 2:630
Constancy, perceptual. See Perceptual constancy
Constituent assembly, 2:618–619
Constraints, analogical, 1:25, 26
Constructional apraxia, 1:48
Construction-integration model, 1:409–410
Constructivism, 2:450
Consumer behavior, 1:321–323
Consummate love, 2:474
Contagion, emotional, 1:354, 2:605
Content externalism, 2:501
Contention scheduling/supervisory attentional system theory, 1:7–8

Contentment, 1:381
Contiguity effect, 2:493–494, 496
Contingency-based learning, 1:245
Contingent aftereffects, 2:479–480
Contingent negative variation, 2:698
Continuity illusion, 1:385–386
Contour (music), 2:549
Contrastive analysis, 1:201

Control
anxiety disorders, 1:42
automaticity, 1:96–97
metacognition, 2:508

Controlled processes, 2:707
Conversation analysis, 1:203

Conversation and dialogue, 1:202–205
inner speech, 1:418–419
multimodal systems, 2:535–538
perspective taking, 2:588–590

Converse, Philip, 2:609
Cooley, Charles Horton, 2:672
Coolidge, Frederick L., 2:792–793
Cooper, Joel, 1:155

Core self-consciousness, 2:674
Correlated personal intentions, 2:444–445
Correspondence bias, 1:84
Correspondence theory of truth, 1:197–198
Cortex, 2:693, 697–698
Cortical blindness, 1:120–121
Cortical deafness, 1:87
Cortical pathways, 1:413
Cortically based affect-processing network, 1:314
Coser, Lewis, 1:434
Cosmides, Leda, 1:229
Costa, Albert, 1:116–117
Counterconditioning, 1:109
Counterfactual thinking, 2:640
Coupled processes, 1:336–338
Courbon, P., 1:364
Covert self-talk, 1:418
Cowan, Nelson, 2:481, 789, 790
Cowey, Alan, 1:227
Craik, Fergus, 1:13, 2:641–642
Cramer, Gabriel, 1:221
Crapse, Trinity, 1:56
Credence, 1:113
Crick, Francis, 1:2, 2:699
Crowds. See Wisdom of crowds effect
Cue learning, 2:714
Cultural relativism, 1:353
Culture
attachment theory, 2:648
child conceptual development, 1:170–171
cognitive dissonance, 1:155
disgust, 1:255
distributed cognition, 1:259
emotion, 1:290–293
knowledge acquisition, 2:451
music, 2:550
Cummins, Robert, 1:200
Cushman, Fiery, 1:329
Custers, Ruud, 1:192
Cutting, James, 2:574
Dallenbach, Karl, 2:481, 483
Damasio, Hanna, 1:317
Dantzig, George, 1:213
Darwin, Charles, 1:314, 349, 352, 2:547, 550, 583
Daum, Moritz, 1:162
Davidson, Donald, 1:4, 5, 32–34, 113, 177–178, 2:500, 510, 597–598
Davidson, Richard, 1:289
Davis, Orin, 2:496
Dawkins, Marion, 1:191
Dawkins, Richard, 1:222
De Clérambault delusion, 1:230
De Finetti, Bruno, 1:264
De Groot, Annette, 1:115
De Morgan's laws of propositional logic, 1:416
De Renzi, Ennio, 1:47
De Sousa, Ronald, 1:432
De Waal, Franz, 1:191
Dead reckoning, 2:714
Deafness
cortical, 1:87
language development, 2:459
pure word, 1:44
Debiasing, 1:207–209, 212
Decay, memory, 2:481
Decay of excitation, 2:706
Deception, 1:345, 354, 2:470–472
Deception, linguistic cues to, 1:210–211
Decision affect theories, 1:21
Decision automation, 1:214
Decision biases, 2:474–475
Decision improvement technologies, 1:211–214.
See also Prediction, clinical versus actuarial
Decision making. See also Heuristics; Judgment
action selection, 2:556–558
affective forecasting, 1:11–13
Allais paradox, 1:19–21
certainty effect, 1:21
contexts, 1:12–13
debiasing, 1:207–209
dissent, 1:257–258
endowment effect, 1:321–323
group, 1:374–375
intuitive vs. deliberative reasoning, 2:746
love, 2:474–475
mental effort, 2:503
models, 1:216–217, 219–220
morality, 2:527
natural language generation, 2:559–600
neuroeconomics, 2:564–567
regret, 2:640–641
rewards, 1:215–216
support for, 1:212
theories, 1:212–213 (see also Decision theory,
philosophical perspectives)
Decision making, neural underpinnings, 1:214–217,
2:564–567
Decision making and reward, computational
perspectives, 1:218–220
Decision rules, 2:600
Decision support systems, 1:211–214
Decision theory, philosophical perspectives, 1:220–224
Decision trees, 1:129
Declarative information, 2:691–692
Declarative memory, 1:130, 224–226, 2:485–487, 766
Declarative/procedural model of language, 1:224–226
Depth perception, 1:232–235. See also Stereopsis
combining depth information, 1:235
multiple-view cues, 1:234–235
non-visual senses, 1:235
pictorial cues, 1:232–234
Depth psychology, 2:629
Descartes, René, 1:37, 151, 178, 188, 194–195, 197, 200, 229, 438–439, 2:447–448, 499, 516, 580, 583, 629, 652, 674, 676

Descriptive thought, 1:239–242. See also Knowledge
by description
Desimone, Robert, 1:63, 2:563
Desirable difficulties perspective on learning, 1:242–243, 251, 2:693, 711

Desire, 1:244–247
explanation of action, 1:331–332, 2:599–600
instrumentalism, 2:634–635
neuroscience perspective, 1:245–246
philosophy and neuroscience integrated, 1:246–247
theories, 1:244–245
DeSteno, David, 1:303
Determinism, 1:329, 361–363, 2:599, 773
Deviant causal chains, 1:144–145, 2:598
Dewar, Michaela, 2:481

Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), 1:8, 9

Dialectic behavior therapy, 1:125
Dialectical decision techniques, 1:258
Dialogue. See Conversation and dialogue
Diana, Princess of Wales, 1:16
Dichotic listening task, 1:315, 2:718

Dictator Game, 1:247–249
Diencephalon, 1:22
Diener, Ed, 1:381
Differentiation, 1:251
Difficult-decision paradigm, 1:154
Diffusion (identity), 2:586
Digit span, 2:788–789
Dijksterhuis, Ap, 2:751, 754
Diminishing marginal utility, 1:221
Directional outcome-motivated thinking, 2:529
Discourse processing, models of, 1:249–250
Discrepancy (reinforcement learning), 2:644–645
Discrete state-space, 1:218

Discrimination learning, training methods, 1:250–252
Disfluencies: comprehension processes, 1:252–253
Disfluency, 2:503–504, 748
Disgust, 1:253–255, 303
Disjunction parameter, 1:416–417

Disjunctive theory of perception, 1:255–256
Disorganized attachment style, 2:648
Dispositionalism, 2:703

Dissent, effects on group decisions, 1:257–258
Dissociation paradigm, 2:724
Dissociations, 1:36, 94, 2:465
Distinction bias, 1:13
Distributed cognitio n, 1:258–260, 335. See also Extended mind
Distributed Cohort Model, 1:157
Disulfiram, 1:10–11
Divided attention and memory, 1:260–263. See also Attention and action; Dual-task experiments; Multitasking and human performance
encoding, 1:261–262
long-term memory, 1:261
perception and working memory, 1:260–261
retrieval, 1:261–263
DNA testing, 1:342
Domain knowledge, 2:796
Domain satisfaction, 1:380
Domain theory, 1:302
Domain-specific self-concepts, 2:669–673
Donald, Merlin, 1:338
Donders, Franciscus, 1:14, 2:630–631, 632
Donnellan, Keith, 1:37, 237
Doppelgängers, 1:100, 101, 102. See also Subjective doubles
Dorsal visual stream, 2:563
Dorsolateral prefrontal cortex, 2:545
Double. See Doppelgängers
Doumas, Leonidas, 1:27
Dove, Guy, 1:174, 175
Down syndrome, 2:459
Down-states, 1:275
Dreams. See Sleep and dreams
Dressing apraxia, 1:48–49
Dretske, Fred, 1:38, 141, 142, 187, 188, 189, 245, 438, 2:703
Drive theory, 1:109–110
Drug abuse, 1:9
Drug classes, 1:8
Drug dependence, 1:9
Dual calculation hypothesis, 2:582
Dual-aspect monism, 1:190, 193
Dualism, 1:194–195, 2:516–518, 583, 639. See also Mind-body problem
Dual-process models
morality, 1:303–304
persuasion, 2:591–592
reasoning, 1:209, 229, 2:745–748
social cognition, 2:706–707
thinking, 2:741–742
Dual-task experiments, 1:162, 260, 2:632, 766–767. See also Divided attention and memory
Duchenne de Boulogne, Guillaume-Benjamin, 1:352
Duchenne smile, 1:282
Duffy, Charles, 2:575
Dumais, Susan, 2:682, 688
Dummett, Michael, 1:177
Duncan, John, 1:53, 63, 74–75, 2:563
Duplex theory of love, 2:473–474
Duration, 2:515, 743–744
Dutch book arguments, 1:264–265
Dworkin, Ronald, 2:526
D’Ydewalle, Géry, 2:127
Dykes, Robert, 2:542
Dynamic programming, 2:477
Dynamic time warping, 2:477–478
Dysexia, 1:267
Dyslexia, acquired, 1:266–267
Dyslexia, developmental, 1:267–269
Dyslexia, phonological processing in, 1:269–271
Dyslexia, types of, 1:266–267. See also Word recognition, visual
Eager learning, 1:129
Ear, 1:85–86, 383
Ebbinghaus, Hermann, 1:400, 2:710, 712
Eberhard, Kathleen, 1:409
Economist, 1:118–120
Ecological approach to visual perception, 2:573
Ecological rationality, 2:557
Economics
Dictator Game, 1:247–249
endowment effect, 1:321–323
intergroup conflict, 1:435
neuroeconomics, 2:564–567
Education
metacognition, 2:508–509
multiple intelligences theory, 2:541–542, 543
spacing effect, 2:712–713
Educational neuroscience, 2:668
Efferent representations, 1:161
Effort expectancy, 2:709
Ego identity, 2:585
Eimer, Martin, 2:612
Eisenhower, Dwight D., 1:213
Ekman, Paul, 1:290, 314, 349, 352, 354
Elaboration likelihood model, 2:591
Elaborative inferences, 1:409
Elaborative rehearsal, 2:641–642
Electrocardiography (ECG), 1:295–296
Electrodermal activity (EDA), 1:295
Electromagnetic neuroimaging methods, 1:56–59
Electromyography (EMG), 1:296, 2:695
Electrooculography (EOG), 2:695
Electrophysiological sleep recordings, 2:695–696
See also Neuroimaging

Eliminative materialism, 1:277–281  
characteristics, 1:277–278  
criticisms of, 1:280  
critiques of folk psychology by, 1:278–280, 358–359  
Eliminativism, 1:329, 2:517, 637–638, 703  
Elizabeth of Bohemia, Princess, 2:499  
Ellis, Albert, 2:106, 2:651  
Ellsberg, Daniel, 2:565  
Ellsberg paradox, 2:565  
Elman, Jeff, 1:157  
Elsner, Birgit, 1:162  
E-mail, 1:204  
Embarrassment, 1:281–282  
Emblems (gesture), 1:372  
Embodied cognition, 1:163, 351  
Embodied conversational agents, 2:536, 537  
Embodiment movement, 1:194, 196–197  
Emergence, 1:283–287  
deducibility, 1:283–284  
epistemological, 1:284, 286–287  
features, 1:180–181  
ontological, 1:284–286  
overview, 1:283  
types, 1:284–287  
Emergent physicalism, 2:601–602  
Emergent properties, 2:512  
Emmons, Robert, 1:324  
Emotion. See also Feelings  
affective forecasting, 1:11–13  
animals, 1:190–193  
attention and, 1:68–73  
bodily expression, 1:292  
cognitive dissonance, 1:154  
computational models, 1:165–166  
confabulation, 1:185  
decision making, 1:21  
dimensional approach, 1:298–300, 314  
discrete approach, 1:297–298, 314  
disgust, 1:253–255  
dreams, 2:699  
embarrassment, 1:281–282  
evolution, 1:190, 2:628  
formal objects, 2:628  
intellectualist accounts, 1:320–321  
intentionality of, 1:428–432  
modal model, 1:311  
music and, 2:550  
perception, 1:431–432  
rationality of, 2:627–630  
regulation of, 2:629  
social contagion, 1:354  
space and, 2:771–772  
transience of, 1:12  
unconscious emotions, 2:749–750  
Emotion, cerebral lateralization, 1:287–289  
Emotion, cultural perspectives, 1:255, 290–293  
Emotion, psychophysiology of, 1:293–297  
amygdala, 1:294  
classical fear conditioning, 1:293–294  
cognitive effects, 1:296–297  
hypothalamus, 1:294–295  
measures, 1:295–296  
Emotion, structural approaches, 1:297–300  
Emotion and moral judgment, 1:300–304.  
See also Moral development  
evidence on, 1:303–304  
guilt, 1:376–377  
moral behavior vs. judgment, 1:300–301  
theories, 1:301–303  
Emotion and psychopathology, 1:304–308  
affective emotional networks, 1:305–308  
morality, 1:303–304  
treatment, 1:305  
Emotion and working memory, 1:308–310  
Emotion processing, 2:665  
Emotion regulation, 1:310–313  
Emotion test batteries, 1:316  
Emotional networks, 1:305–308  
Emotional recognition, neuropsychology of, 1:313–319  
developmental processes, 1:319  
facial channel, 1:317–318  
gender, 1:319  
lexical channel, 1:318–319  
methodology, 1:315–316  
neuroimaging studies, 1:317–319  
prosodic channel, 1:318  
theories and models, 1:314–315  
Emotional Stroop effect, 2:723  
See also Intentionality of emotion  
animals, 1:189–193  
unconscious emotions, 2:749–750  
Empathy, 1:301, 2:521  
Empiricism  
knowledge acquisition, 2:449–450  
language development, 2:455  
Empty love, 2:474  
Endogenous attention, 1:68  
Endophenotypes, 1:306  
Endowment effect, 1:321–323  
Energetic masking, 1:385  
Engelbart, Douglas, 1:52, 213  
English language
Black English Vernacular, 1:118–120
substandard vs. nonstandard, 1:119
Engram, 2:487
Entorhinal cortex, 2:484
Entrainment, 1:203
Environmentalism, 1:335
Envy, 1:323–324, 2:658–659. See also Jealousy; Resentment
Epigenetic factors, 1:420
Epilepsy, 1:94, 101, 231, 2:484
Epiphenomenalism, 1:197, 2:500, 517, 602
Episodic buffer, 2:790, 792
Episodic memory, 1:130, 144, 402, 2:491–492
Epistemic action, 1:336
Epistemology
anti-individualism, 1:39
belief and judgment, 1:112–113
Epithalamus, 1:22
Epstein, William, 2:581
Ericsson, K. A., 1:63, 439
Erikson, Erik, 2:585–586, 647
Eros, 2:473
Errorless learning, 2:693
Escher, M. C., 1:233
Estes, Zachary, 1:180
Ether Day, 1:30
Ethnic competition theory, 1:435
Ethnocentrism, 1:435
Ethnography, 1:259
Euclid, 2:580, 720
Evans, Gareth, 1:37, 38, 2:453, 570
Evans, Jonathan, 2:740–741, 748
Event memory, development, 1:325–326
Event-related potentials (ERPs), 1:31, 57–59, 253, 275, 316, 2:696, 698
Events, actions as, 1:4–5, 2:597
Evidence accumulation models, 2:633
Evolution
attention, 1:53–56
attraction, 1:82
brain/mind, 1:190–193
cognition, 2:792–793
emotion, 1:190, 2:628
envy, 1:323
face perception, 1:345, 352
hearing, 1:382
jealousy, 2:442
love, 2:474–475
music, 2:547–548, 550
sleep, 2:695
teleology, 2:737
working memory, 2:791–793
Excitation pattern (hearing), 1:384
Executive attention, 2:504
Executive control, 2:502–503
Executive functions, 1:91
Exemplars, 1:135, 139, 173
Exercise and the brain, 1:326–328
Expectancy theory, 2:605–606
Expectation, rationality and, 2:629
Expected utility theory, 1:19–20, 2:565
Experience
explanatory gap, 1:333–335
language development, 2:455
olfactory, 2:700–701
qualia, 2:635–636
teleology, 2:737–739
theory of appearing, 2:737–739
Experiential memories, 2:571–572
Experimental philosophy, 1:328–330
Explanation of action, 1:330–333
causal, 1:332
characteristic features, 1:330–332
instrumental, 1:333
teleological, 1:332–333
teleology, 2:735–737
Explanatory gap, 1:333–335, 2:602–603
Explicit memory, 1:400–402, 2:766
Expressivism, 1:302
Extended mind, 1:335–338. See also Distributed cognition
criticisms of, 1:337–338
original thesis, 1:336–337
other formulations, 1:338
overview, 1:335–336
Extended self-consciousness, 2:674
Extension. See Intension and extension
External focus, 2:531–532, 693
Externalism, 1:152, 336. See also Vehicle externalism
Extreme-groups studies, 1:423, 2:795
Eye gaze, 2:537
Eye movements during reading, 1:338–340
Eyewitness memory, 1:340–342
confidence, 1:341–342
event testimony, 1:340–341
identification testimony, 1:341
miscarriages of justice, 1:342
Eysenck, Hans, 1:14, 105
F scale, 2:609
Face cells, 1:345
Face perception, 1:343–347
development, 1:344–345
impaired, 1:346–347
information gathering, 1:345
neural mechanisms, 1:345–346
recognition, 1:343–344

F scale, 2:609
Face cells, 1:345
Face perception, 1:343–347
development, 1:344–345
impaired, 1:346–347
information gathering, 1:345
neural mechanisms, 1:345–346
recognition, 1:343–344
Face recognition vendor test, 1:348–349
Face space model, 1:347–348
Facial action coding system, 1:352–353
Facial expressions, computational perspectives, 1:349–351
Facial expressions, emotional, 1:317–318, 351–354
contagion, 1:354
control of, 1:353–354
current theories, 1:353
gender, 1:354
history of study, 1:352–353
meaning, 1:351–352
Facial feedback, 1:354
Facoetti, Andrea, 1:268
Fail, G., 1:364
Fajen, Brett, 2:575
False belief task, 1:359
False cognates, 1:115
Familiarity, 1:24
Fara, Delia Graff, 1:237–238, 239
Farrell, Simon, 2:686
Fatuous love, 2:474
Fazio, Russell, 1:81, 155
Fear
anxiety disorders, 1:40
classical conditioning, 1:293–294
facial expressions, 1:317
innate and learned, 1:294
Featural similarity, 1:25–26, 2:688–690
Feature integration theory, 2:562–563, 764
Feature matching, 1:234
Feature overlap model of semantic memory, 2:680–681
Feature search, 2:562
Fechner, Gustav, 2:549, 743
Feedback
frequency, 2:531–532
mnemonic benefits, 2:661, 712
normative, 2:532
positive, 2:532
self-controlled, 2:532
Feedback-based learning, 1:393, 2:531–532
Feeling of another's knowing, 1:252
Feeling of knowing, 1:252
Feeling theories, 1:430
Feelings, 1:244–245, 299, 431–432. See also Emotion
Fehr, Beverley, 2:474
Fehr, Ernst, 2:567
Feldman, Jacob, 1:173
Feltz, Adam, 1:329
Fernandes, Myra, 1:262, 263
Fernandez-Duque, Diego, 1:60
Festinger, Leon, 1:153–155
Fiala, Brian, 1:328
Field, David, 2:575
Fillers, in speech, 1:252–253
First fixation duration, 1:339
First-person authority, 1:39
Fischer, Kurt, 2:669, 671
Fishbein, Martin, 1:79, 80
Fisher, Ronald, 1:389
Fixations (vision), 1:338–339
Fixed word order languages, 2:618
Flashbulb memories, 1:16
Flat maximum principle, 2:614
Flavor, 2:732–734
Flege, James, 1:117
Flicker technique, 1:145
Flickers of freedom, 1:363
Fluctuation theories, of spacing effect, 2:711
Flynn, James R., 1:355–357
Flynn effect: rising intelligence scores, 1:355–357, 390
Focalism, 1:12
Focus of expansion, 2:573–576
Folk psychology, 1:358–360
alternative theories, 1:359–360
critiques of, 1:278–280, 358–359, 2:634–635
mind-body problem, 2:517
overview, 1:358
simulation theory, 1:359
text theory, 1:358–359
Folstein, Susan, 1:92
Food and Drug Administration (FDA), 1:10, 93, 289
Foot-in-the-door effect, 2:594
Foreclosure (identity), 2:586
Forrester, Jay, 1:213
Forsythe, Robert, 2:688
Forsythe, Robert, 2:688
Forward masking, 2:760, 761
Foundationalism, 2:489–490
Fovea, 1:338
Fox, Nathan, 1:289
Fragopanagos, Nickolaos, 1:73
Framework of relevance theory, 2:511
Francis, Martha, 1:210
Franck, Julie, 2:462
Frankfurt, Harry, 1:362, 2:599
Frankfurt-style examples, 1:362–363
Fraser, Scott, 2:594
See also Social loafing
Free sorting, 1:393
See also Freedom of action; Voluntary action, illusion of
Free word order languages, 2:618
Freedman, Jonathan, 2:594
Freedom
action and, 2:599
crossword, 1:361
debate about, 1:361
flickers of, 1:363
threats to, 1:361–362
Freedom of action, 1:360–363
  automatism, 1:93–95
debate about, 1:361
illusion of, 2:772–773
Free-field viewing procedure, 1:315
Frege, Gottlieb, 1:424, 2:570
Fregoli, Leopoldo, 1:364
Fregoli delusion, 1:364–365
Frequency, of sound, 1:89
Frequency effects in word recognition, 1:366–367
  Freud, Sigmund, 1:109, 188, 279, 394, 2:513, 609, 610, 629, 647, 651, 674, 676, 698–699, 739, 749, 751
Frication, 2:717
Fridlund, Alan, 1:353
Friendship, 2:593, 648–649
Friesen, Wallace, 1:314, 352
Frijda, Nico, 1:353
Friston, K., 1:64
Frith, Chris, 1:231
Frontal eye fields (FEF), 1:58, 413
Frontal lobes, 1:18, 22, 26
Frontal-parietal network, 1:58–59
Fulcher, Eamon, 1:72
Function assignment, 2:618–619
Functional elements of language, 2:608
Functional reduction, 1:283–284
Functionalism
  belief, 1:113
  consciousness and embodiment, 1:195
  extended mind, 1:337
  mind-body problem, 2:517–519
  realizer functionalism, 2:518–519
  role functionalism, 2:518–519
Functions, 1:138
Fundamental attribution error, 1:84
Furman, Wyndol, 2:649
Furnas, George, 2:682
Gabor filters, 1:350
Gagne, Christina, 1:180
Gainotti, Guido, 1:288, 314
Galton, Francis, 2:660, 726
Game theory, 1:223, 436, 2:565, 566
Gamma activity, 1:275
Ganong effect, 2:718
Gardner, Howard, 2:541–543
Garety, Philippa, 1:231
Gaskell, Gareth, 1:157
Gauthier, David, 1:223
Gaze duration, 1:339
Gaze-contingent moving window paradigm, 1:339
Geach, Peter, 1:112
Geertz, Clifford, 1:259
Gender
  child stereotyping, 1:170
  emotional recognition, 1:319
  facial expressions, 1:354
  jealousy, 2:442
  memory, 1:326
  sex vs., 1:369
Gender differences in language and language use, 1:369–370
  motivations, 1:370
  patterns, 1:370
  sex-determined, 1:369
  social meaning, 1:369–370
General intelligence (g), 1:421–423, 2:541
Generalized anxiety disorder, 1:41
Generation effect (memory), 1:402
Generative coding, 1:64
Genes and linguistic tone, 1:371–372
Genetic algorithms, 2:557
Genetic factors
  anxiety, 1:42
  attitudes, 1:77
  autism, 1:92
  dyslexia, 1:269
  face perception impairments, 1:347
  heritability, 1:389–390
  intelligence, 1:419–420
  language, 1:413–417
  musical capacities, 2:547
  schizophrenia, 2:664
  substance use disorders, 1:10
  Williams syndrome, 2:775–776
Genotype, 2:775
Gentner, Dedre, 1:25, 27, 28, 2:512, 690
Geometric models of similarity, 1:687–688, 690
Gesture and language processing, 1:372–374
  child development, 2:456
  speaking, 1:373
  thinking, 1:373
  types of gesture, 1:372–373
  variation in gesture, 1:373–374
Gettier, Edmund, 1:112
Gibson, J., 1:162, 2:573, 575–576
Gilbert, Margaret, 2:445
Glenberg, Arthur, 1:163, 174
Glimcher, Paul, 2:566–567
Global aphasia, 1:44, 279
Global self-esteem, 2:669, 671–673
Glucksberg, Sam, 1:180, 2:511, 512
Goal direction, 2:736
God
  Berkeley’s idealism, 1:397–398
  Leibniz and mind-body relation, 2:499
Goldberg, Robert, 1:328
Goldberg, Sanford, 1:38
Golden rule of predictive modeling, 2:613–614
Goldie, Peter, 1:431
Goldman, Alvin, 1:229, 359, 438
Goldstein, Karl, 2:513
Gollan, Tamar, 1:117–118
Go/no go tasks, 2:632
Goodwin, Charles, 1:203
Gopher, Daniel, 1:61
Gopnik, Alison, 1:173, 359, 2:677
Gordon, Robert, 1:359
Goro, Takuya, 1:416
Gorry, Anthony, 1:214
Goswami, Usha, 1:27, 268
Gottardo, Alexandra, 2:796
Graesser, Arthur, 1:250
Grammar
  agreement, 2:461–463, 618–619, 727–729
  child development, 2:458
  generating structure, 2:463–464
  musical, 2:550
  universal, 1:413–414, 417, 2:458
Grandin, Temple, 1:191
Gratiosolet, Pierre, 1:352
Gray, Jeffrey, 1:315
Green, David, 1:116
Greene, Joshua, 1:303, 329, 2:747
Greene, Robert, 2:642
Greenwald, Anthony, 2:753
Grent-’r’Jong, Timeke, 1:59
Grice, H. P., 2:510
Griffin, Don, 1:190
Griggs, Richard, 1:227
Grigorenko, Elena, 1:269
Grossberg, Stephen, 2:480, 563
Grounding, in conversation, 1:204
Group action, 1:159
Group decision making, 1:374–375
  benefits, 1:257, 374, 2:777
  dissent, 1:257–258
  limitations, 1:257–258, 374–375, 2:777–778
  motivational effects, 1:375
  social comparison processes, 1:374–375
  wisdom of crowds effect, 2:776–778
Group differences, in working memory, 2:790–791
Group identification, 1:433
Group perception, 1:432–433
Group polarization phenomenon, 1:374
Group threat, 1:433
Groups. See Intergroup conflict; Intergroup conflict, models of; Wisdom of crowds effect
Grove, William, 2:613
Guidance hypothesis, 2:531
Guided search, 2:764–765
GUILD, 1:7
Habits, 1:7
Habituation, 1:412
Haggard, Patrick, 2:612, 773
Hahn, Ulrike, 2:690
Haidt, Jonathan, 1:302, 303, 2:610
Hair cells, 1:85–86, 383
Hájek, Alan, 1:222
Halford, Graeme, 1:27
Hallucinations, 1:99–100, 255–256, 2:554
Hampton, James, 1:173, 180
Happiness, 1:379–382. See also Unhappiness
  concept of, 1:379–381
  conditions for, 1:381
  measurement of, 1:381
  prevalence of, 1:381
  promotion of, 1:381–382
  religion and, 2:650–651
Hariri, Ahmad, 1:318
Harley, Trevor, 1:73
Harman, Gilbert, 1:227
Harmon-Jones, Eddie, 1:155, 289
Harmony, 2:549
Harshman, Richard, 2:682
Harter, Susan, 2:670–673
Harvard Twin Study of Substance Abuse, 1:10
Hasher, Lynn, 1:18
Hassin, Ran, 2:753
Hauser, Marc, 1:302
Hayes, Steven, 1:107
Haynes, John-Dylan, 2:773
Heading (motion), 2:573–574
Heal, Jane, 1:359
Health
  happiness, 1:379
  hypochondria, 1:394–395
Hearing, 1:382–387. See also Deafness
  auditory masking, 1:88–90
  auditory scene analysis, 1:386
constancy, 1:386–387
contents of, 1:387
depth perception, 1:235
information conveyed by, 1:383–384
masking, 1:384–385
neural basis of, 1:85–88
objects of, 1:387
peripheral vs. central processes, 1:384
phenomenology of, 1:388
pitch, 1:388
place vs. temporal information, 1:384
space, 1:387–388
speech, 1:388–389
time, 1:388
truth vs. illusion, 1:385–386
unconscious perception, 2:754
word recognition, 2:783–784
Hearing, philosophical perspectives, 1:387–389
Heart rate, 1:295–296
Heautoscopy, 1:99–102
Hebb, Donald, 2:488
Hebbian learning, 1:279, 2:686
Hedonic adaptation, 1:12
Hedonic level of affect, 1:381
Hegel, Georg Wilhelm Friedrich, 1:198
Heider, Fritz, 1:83
Heilmann, Kenneth, 1:49, 314
Helmholtz, Hermann von, 2:549, 752
Helwig, Charles, 2:526
Hemianopia, 1:120–121
Hemianopic autoscopy, 1:100
Hemiplegia, 1:34–36
Hemodynamic neuroimaging methods, 1:57–59
Hendrick, Clyde, 2:473
Hendrick, Susan, 2:473
Herbert, Martha, 1:91
Heritability, 1:389–390
Heritage, John, 1:203
Heritage language and second language learning, 1:391–392
bilingual processing, 1:114–115
cognitive benefits of bilingualism, 1:115–118
Hermann, Rudolph, 1:161
Hernández, Mireia, 1:116
Herpes simplex encephalitis (HSE), 1:22–23
Herrnstein, Richard, 1:167, 218, 355
Hertwig, Ralph, 2:777
Herzog, Stefan, 2:777
Heteronomy, 2:525–526
Heuristic systematic model, 2:591–592
Heuristics. See also Decision making
availability heuristic, 1:102–103
common ground determination, 2:616–617
persuasion, 2:592–594
representativeness heuristic, 2:657–658
Heuristics and biases framework, 2:637
Hewes, Gordon, 2:547
Hick-Hyman law of mean, 2:632
Hicks, Jason, 1:263
Hicks, Robert, 1:261
Hidden Markov models, 2:478
High throughput experiments, 2:668
High-dimensional vector coding, 1:279
Hill, Sarah, 1:323
Hillyard, Steve, 1:58
Hinde, Robert, 2:646
Hippocampus, 1:24, 133, 2:484–485, 488
Hirsch, Joy, 2:721
Hirshleifer, Jack, 1:376
Histrionic personality disorder, 2:555
Hitch, Graham, 1:15, 261, 2:686, 766, 786, 789, 791
Hobbes, Thomas, 1:187, 2:600, 608–609
Hobson, Allan, 2:699
Hofstadter, Douglas, 1:223, 2:674
Holism, 1:198
Holmes, Oliver Wendell, Jr., 2:469
Holway, Alfred, 2:581
Holyoak, Keith, 1:228, 229
Hommel, Bernhard, 1:161, 162
Homo erectus, 2:792
Honesty, 2:527
Hopfinger, Joseph, 1:58, 59
Horgan, Terence, 2:635
Horn, Robert, 1:51
Hornickel, Jane, 2:542
Horopter, 2:719–720
Horowitz, Joel L., 2:721
Hratchian, 1:387
Hruschka, Jack, 1:376
Husserl, Edmund, 2:726
Hu, Xiangen, 3:198
Hubbard, Edward, 2:726
Huber, Robert, 1:193
Hughes, Gethin, 2:612
Huijbers, Willem, 1:262
Hull, Clark, 1:108, 109–110, 111
Human classification learning, 1:392–394. See also
Category learning, computational perspectives;
entries beginning with Categorization
Human Genome Project, 1:420
Human-computer interaction
collection, 1:204
distributed cognition, 1:258
emotion, 1:73, 165
Hume, David, 1:151, 302, 2:523, 583, 629, 652
Hummel, John, 1:28
Huntington's disease, 1:132, 2:486, 558
Hurley, Susan, 1:335
Husserl, Edmund, 2:674
Identity theory, 1:302
Hutcheson, Francis, 1:302
Hutchins, Edwin, 1:258, 259, 335
Hybrid theories, of working memory, 2:790
Hyman, John, 1:5
Hyperintension, 1:426
Hyperspace analogue to language model, 2:682
Hypochondria, 1:394–395
Hypocrisy, 12
Imitation, 1:196
Implicit memory, 2:516–638. See also Mind-brain identity theory; Type identity theory
Idéalisme, 1:397–399
Berkeley, 1:397–398
consciousness and embodiment, 1:198
materialism vs., 1:397–399
panpsychism, 1:399
phenomenalism, 1:398–399
reductivism, 2:637
Ideational apraxia (IA), 1:47–48
Identification disorder, 1:365
Identity achievement, 2:586
Identity aftereffect, 1:344
Identity and identification. See also Personal identity;
Personal identity, development of
group identification, 1:433
social identity, 1:435
Identity theory, 2:516, 638. See also Mind-brain identity theory; Type identity theory
Ideology, 2:609–611
Ideomotor action, 1:162
Ideomotor apraxia (IMA), 1:47–48
Ideomotor theory, 1:161–162, 2:772
Imagery. See Visual imagery
Imitation, 1:162
Immediate memory, 1:14–15, 2:766, 787
Immigrants, languages of, 1:391–392
Immune neglect, 1:12
Immunotherapies, 1:11
I-mode, 1:159–160
Impact bias, 1:12
Implicit memory, 1:400–403
explicit memory vs., 1:401–402
modern research, 1:400–401
neuroimaging studies, 1:402–403
unconscious perception, 2:754
Implicit perception, 2:577, 751. See also Unconscious perception
Unconscious perception
In vitro studies, 2:488
Inattentive amnesia, 2:579
Inattentual blindness, 1:403–406
consciousness, 1:66–67, 2:577–579
factors influencing, 1:404–405
nature of, 1:403–404
related phenomena, 1:405–406
visual search vs., 1:404
Incentive salience, 1:10
Incentive-sensitization model, 1:10
Incompatibilism, 1:329, 361–363, 2:599
Incongruent behavior, 2:704–705
Independence axiom, 1:20
Indeterminism, 1:363
Indexical concepts, 2:603
Indexical thought, 1:406–407
Individual differences, in working memory, 2:790–791
Individualism, 1:291
Individualism about cognition, 1:37
Induced compliance paradigm, 1:154
Inductive reasoning, 1:226
Inertial navigation, 2:714
Infatuated love, 2:473–474
Inference, 1:26, 2:677, 796
Inferences in language comprehension, 1:408–410
bridging inferences, 1:408–409
elaborative inferences, 1:409
factors influencing, 1:410
level of language representation, 1:408
predictive inferences, 1:409
process model, 1:409–410
Inferential theories of concepts, 1:172–173
Inferotemporal cortex, 1:133
Infinite utilities, 1:222–223
Information processing
knowledge acquisition, 2:450
limitations, 1:74–75
models, 1:64, 422
neuroscience evidence, 2:542–543
schizophrenia, 2:665–666
sleep, 2:696–698
visual search, 2:765
Information processing speed, 1:13–19
Informational masking, 1:385
Information-integration category learning, 1:132
Inhibition of return, 1:410–413
characteristics, 1:410–411
consequences, 1:412
mechanisms underlying, 1:411–412
neural correlates, 1:412–413
overview, 1:410
purposes, 1:411
Inhibitory control model, 1:116
Initial state probability vector, 2:478
Innate concepts, 1:170
Innateness and parameter setting, 1:413–417
Jensen, Art, 1:14
Jespersen, Otto, 2:548
Johnson, Mark, 1:60
Johnson, Samuel, 1:399
Johnson-Laird, Philip, 1:228, 2:740
Joint action, 1:159
Joint commitment, 2:444–445
Joint or collective intention, 2:443–445.
See also Collective action
Jolicoeur, Pierre, 1:62
Jones, Mary Cover, 1:109
Jost, John, 2:610
Judgment. See also Belief and judgment;
Decision making; Heuristics
anchoring, 1:28–29
availability heuristic, 1:102–103
debiasing, 1:207–209
moral, 1:300–304
two-system model of reasoning, 2:747
Jung, Carl Gustav, 2:698
Just, Marcel, 2:794
Just noticeable difference (jnd), 2:743
Justification, 2:490–491
Justified belief, 1:229–230
Kahana, Michael, 2:496
Kahneman, Daniel, 1:29, 60, 102, 212, 222, 260, 2:502, 565, 566, 629, 657–658
Kant, Immanuel, 1:302, 2:514, 526, 652, 674
Kaplan, David, 1:407
Kapur, Shitij, 1:231
Kastner, Sabine, 1:59
K-complexes (KCs), 2:697–698
Kelley, Harold, 2:587
Kemper, Tom, 1:91
Kennedy, John F., 1:16
Kenrick, Douglas, 2:474
Kentridge, Robert, 2:579
Kernberg, Otto, 1:124
Keni, Wilson Bryan, 2:724, 753
Kieras, David, 1:62
Kihlstrom, John, 2:751, 752, 753
Killen, Melanie, 2:526
Kim, Jaewon, 1:33, 283
Kinderman, Peter, 1:231
Kinsbourne, Marcel, 1:261, 289
Kintsch, Walter, 1:63, 249, 409
Kirby, Simon, 2:548
Kirsch, David, 1:336
Klein, Raymond, 1:411
Kleist, Karl, 1:48
Kleitman, Nathaniel, 2:695
Knafo, Ariel, 1:248
Knobe, Joshua, 1:329
Knoblich, Günther, 1:162
Know-how, philosophical perspectives, 2:447–448
Knowledge. See also Self-knowledge
memory, 2:489–492
metacognitive, 2:508
prior, 2:524
semantic, 2:680–683
semantic dementia, 2:678–679
Knowledge acquisition in development, 1:170, 2:449–451. See also Concepts, development of;
Language development
content and structure, 2:450–451
context and culture, 2:451
process, 2:449–450
Knowledge acquisition, skill learning vs., 2:691
Knowledge by acquaintance, 2:452–453
Knowledge by description, 2:452. See also
Descriptive thought
Knowledge-how, 2:447
Knowledge-that, 2:447
Koch, Christof, 1:2
Koenig, Harold, 2:651
Koenigs, Michael, 1:304
Kohlberg, Lawrence, 2:526
Kolmogorov, Andrey, 1:265
Kornblum, Sylvan, 1:162
Korsakoff’s amnesia, 1:183, 184, 400
Kosfeld, Michael, 2:567
Kosslyn, Stephen, 2:758
Kraepelin, Emil, 1:304
Kraljic, Tanya, 2:621
Kraus, Nina, 2:542
Kripke, Saul, 1:37, 111, 152, 177, 237
Kroll, Judith, 1:115
Kuhn, Thomas, 2:668
Kunde, Wilfried, 1:162
Laming, David, 2:642
Lamme, Victor, 2:578
Landauer, Thomas, 2:682, 688
Landis, John, 1:148
Lane keeping, 2:576
Langdell, Christopher Columbus, 2:469
Langdon, Robyn, 1:231
Language. See also Aphasia; entries on Bilingualism
autism, 1:90
bilingual processing, 1:114–115
biological basis, 2:455–456
cognition, 1:278–279
concepts, 1:169, 176–179
deception, 1:210–211
declarative/procedural model, 1:224–226
gender differences, 1:369–370
Index
819

music in relation to, 2:547–548
natural language generation, 2:558–560
philosophy of language, 1:424–426
Language comprehension
inferences, 1:408–410
working memory, 2:794–796
Language development, 2:455–459. See also
Knowledge acquisition in development
atypical, 2:459
bilingual, 2:458–459
biological basis, 2:455–456
genetic factors, 1:371–372
historical and theoretical perspectives, 2:455
innateness and parameter setting, 1:413–417
lexicon, 2:457–458
phonology, 2:456–457
pragmatics, 2:456
Language development, overregulation in, 2:460–461
Language learning, 1:250–251
Language of thought, 2:652–653
Language processing
bilingualism, 1:115–118
compound words, 1:163–165
discourse processing, 1:249–250
gesture, 1:372–374
perspective taking, 2:588–590
Stroop effect, 2:722
working memory, 2:793–796
Language production. See Planning in language production; Production of language
Language production, agreement in, 2:461–463.
See also Production of language; Syntactic production, agreement in
Language production, incremental processing in, 2:463–464. See also Production of language
Latent learning, 1:242
Latent semantic analysis, 2:682, 688
Lateralized readiness potential, 2:612
Lau, Hakwan, 1:123
Laurence, Stephen, 1:53, 172
Lavie, Nilli, 2:578
Law. See entries beginning with Legal
Layered control architectures, 2:464–466
Leahy, Thomas Hardy, 1:111
Leaky competing accumulator, 2:558
Learned helplessness, 1:289
Learning
associative learning, 1:78, 170
attitude change, 1:78
bias, 1:139
category learning, 1:130, 138–140, 392–394
child development, 2:449–451
concepts, 1:170
desirable difficulties, 1:242–243, 251, 2:693
desire and, 1:245
discrimination learning, 1:250–252
errorless learning, 2:693
folk psychology view of, 1:279
latent, 1:242
machine learning, 2:600
metacognition, 2:508–509
modeling of causal learning, 2:523–525
molecular mechanisms, 2:488–489
motor learning, 2:529–532
overlearning, 1:242
performance vs., 1:242–243, 251
process of, 1:243
self-efficacy, 2:509
semantic dementia, 2:679
skill learning, 2:486, 691–694
sleep and, 1:251–252
spacings effect, 2:712–713
transfer, 1:242–243, 251, 2:692–693
vocal, 2:547
Learning styles, 2:467–468. See also Multiple intelligences theory
LeDoux, Joseph, 1:191, 294
Lee, David, 2:575
Lee, John Alan, 2:473
Leech, Robert, 1:27
Left hemisphere of brain
apraxia, 1:46–49
attention, 1:58
confabulation, 1:185
emotion, 1:287–289, 314–315
language processing, 2:456, 718
Left parietal cortical damage, 1:8, 46
Legal formalism, 2:469
Legal realism, 2:469
Legal reasoning, psychological perspectives, 2:468–470
automatism as defense, 1:94–95
deductive vs. analogical approaches, 2:468
formalism vs. realism, 2:469
scientific reasoning vs., 2:469–470
Leibniz, Gottfried Wilhelm, 1:199, 2:499
Leonardo da Vinci, 2:580
Leont’ev, Alexei, 1:259
Levent, William, 2:464
Levine, Joseph, 1:333
Levinson, B. W., 2:456–459
Levy, Jerre, 1:315
Lewandowsky, Stephen, 2:642, 685, 686
Lewis, David, 1:424, 2:448
Lexical development, 2:457–458, 778–779, 785
Lexical selection and retrieval, 2:617–618
Li, Li, 2:574, 575, 576
Libben, Gary, 1:164
Liberman, Alvin, 1:162
Libertarians, 1:361, 363
Libet, Benjamin, 2:611–612, 773
Lie detection, 2:470–472
control question technique, 2:470–471
effectiveness and value, 2:471–472
relevant-relevant technique, 2:470
scientific foundation, 2:471–472
screening techniques, 2:471
Liepmann, Hugo, 1:47–49
Life ability of the person, 1:379–380
Life satisfaction, 1:380–381
Lightness constancy, 2:580
Liking, 2:473, 593
Limen, 2:724, 751
Limited-capacity systems, 2:787
Lineup procedures, 1:341–342
Linguistic determinism, 1:178–179
Linguistic Inquiry and Word Count (LIWC, computer program), 1:210–211
Linguistic theory, 1:413–414, 424–426
Linguistic tone, 1:371–372
Linguistics, 1:45
Linkwords, 2:522–523
LISA (learning and inference with schemas and analogies), 1:28
Livability of the environment, 1:379
Livengood, Jonathan, 1:329
Llewellyn, Karl, 2:469
Llinas, Adolfo, 2:698
Locality effect, 2:685
Location illusion, 2:733
Loci mnemonic strategy, 2:522
Locke, John, 1:151, 194–195, 200, 397–398, 2:652, 674, 703
Locomotion, prospective control of, 2:533–534
Loewenstein, George, 2:566
Loftus, Elizabeth, 1:340
Logan, Gordon, 1:98
Logic, 1:227
Logical behaviorism, 2:517
Logie, Robert, 1:261, 2:790
Lombroso, Tania, 1:329
Lomo, Terje, 2:643
Long, Debra, 2:796
Long, Michael, 1:391
Long-term memory, 1:16, 261, 2:766
Long-term potentiation, 2:488, 645
Looking, prospective control of, 2:533
Looking-glass self, 2:672
Loomis, Alfred, 2:698
Loomis, Jack, 2:576
Loss aversion, 1:322
Lovaas, O. Ivar, 1:93
Love, 2:473–475
attraction, 1:82–83
relationship development, 2:649–650
research applications, 2:475
theories, 2:473–475
types and styles, 2:473
Luce, R. Duncan, 1:218, 2:633
Lucid dreaming, 2:699
Ludlow, Peter, 1:237, 238
Ludus, 2:473
Luria, Alexander, 1:259, 2:513
Lycan, William, 1:371–372
MacDonald, William, 2:482
Machine learning, 2:600
Machine speech recognition, 2:477–478
Mack, Arien, 1:404, 2:577, 579
Mackintosh, Bundy, 1:72
MacLean, Paul, 1:314
Maglio, Paul, 1:336
Magnetic resonance imaging (MRI), 2:664
Magnetencephalography (MEG), 1:57, 70, 75, 273–277, 316, 2:696
Maher, Brendan, 1:128, 231
Mahon, Bradford, 1:174, 175
Maibom, Heidi, 1:358, 360
Maintenance rehearsal, 2:641–642
Malinowski, Bronislaw, 1:259
Mallon, Ron, 1:329, 2:634
Management decision system, 1:213
Mandarin, 1:417
Mannia, 2:473
Manner of articulation, 2:717
Mannes, Albert, 2:778
Maps and mapping
analogical, 1:26
argument, 1:50–52
cognitive maps, 2:771
Marcel, Anthony (Tony), 2:725, 754
Marcia, James, 2:586
Margolis, Eric, 1:53, 171, 172
Markov assumption, 2:524
Marr, David, 2:487
Marsh, Richard, 1:263, 2:752
Marslen-Wilson, William, 1:156, 157, 2:784
Martin, Alex, 1:174, 175
Martin, C. B., 1:398
Martin, M.G.F., 1:256
Martínez, Antigone, 1:58
Martínez-Manrique, Fernando, 1:174
Marx, Karl, 1:434
Masking, 1:97
auditory, 1:88–90, 384–385
backward, 1:123, 2:753, 760, 761
energetic, 1:385
forward, 2:760, 761
informational, 1:385
pattern, 2:754, 760
sandwich, 2:754
simultaneous, 2:760
typology, 2:760
uses, 2:760, 762
visual, 2:760
Masolino, 1:233
Mass hysteria, 2:605
Matching (gesture), 1:373
Matching bias, 2:740–741
Materialism. See also Physicalism
eliminative, 1:277–281
explanatory gap, 1:335
idealism vs., 1:397–399
Mathews, Andrew, 1:72
Matthews, Gerald, 1:73
Maupassant, Guy de, 1:102
Maximin strategy, 1:221
Maximin strategy, 1:221
Maylor, Elizabeth, 1:17
McCabe, Kevin, 2:566–567
McClelland, Jay, 1:157, 2:563, 646, 785
McCollough, Celeste, 2:479, 480
McCollough effect, 2:478–480
McCorkie, George, 1:339
McCullough, Michael, 1:324
McDowell, John, 1:37, 38, 2:570
McGeoch, John, 2:482
McGurk effect, 1:388, 2:718
McKee, Suzanne, 2:582
Mead, Margaret, 1:259, 352
Meatus, 1:383
Mechelli, Andrea, 1:117
Mechsner, Franz, 1:162
Medial prefrontal cortex, 2:675
Medial temporal lobe (MTL), 1:22–24, 2:484–485, 721
Medin, Douglas, 1:173
Meditation, 1:107
Meehl, Paul, 2:613
Melancholia, 1:394
Melody, 2:549–550
Melokinetic apraxia (MKA), 1:48
Meltzoff, Andrew, 1:359
Memory. See also Semantic memory; Sequential memory, computational perspectives; Serial memory; Working memory
accuracy, 1:16, 340–342, 2:480–483, 492
activated memory, 2:788
aging, 1:13–18
amnesia, 1:22–25
anesthesia, 1:31
associative memory, 2:487–488
autobiographical memory, 1:24, 2:678, 679, 759
category learning, 1:130
causal theories, 1:144–145
confabulation, 1:183–184
consolidation, 2:481–482, 485
declarative memory, 1:130, 224–226, 2:485–487, 766
discourse processing, 1:249–250
divided attention, 1:260–263
encoding, 1:23, 243, 261–262, 401–402, 2:661
episodic memory, 1:130, 144, 402, 2:491–492
event memory, 1:325–326
experiential memories, 2:571–572
explicit memory, 1:400–402, 2:766
eyewitness memory, 1:340–342
folk psychology view of, 1:279
iconic memory, 1:14–15
immediate memory, 1:14–15, 2:766, 787
implicit memory, 1:400–403, 2:754
information processing, 1:13–18
long-term memory, 1:16, 261, 2:766
mnemonic strategies, 2:522–523
nondeclarative memory, 1:24, 130
objects of, 2:571–572
overview, 2:765–766
perception and, 2:752
perceptual memory, 2:767
perceptual representation, 1:133, 402
procedural memory, 1:130, 132, 224–226, 402
propositional memories, 2:571–572
rationality and, 2:629
recognition, 1:16
rehearsal, 2:641–643
reminiscing style, 1:326
retrieval, 1:23–24, 243, 261–262, 2:482, 492–496, 523, 660–661
sensory memory, 2:788
short-term memory, 1:182, 2:766, 787–788
source memory, 1:16
spatial effect on learning, 2:710–713
storage, 1:23–24, 243, 422, 2:485
testing effect, 2:660–661
traces, 1:144–145
visual imagery, 2:759
visual memory, 2:766–767
Memory, interference with, 2:480–483
eyewitness memory, 1:340–342
Stroop effect, 2:722–723
Memory, neural basis, 1:402–403, 2:483–489. See also
Categorization, neural basis
associative memory, 2:487–488
declarative/nondeclarative distinction, 2:485–487
future directions, 2:489
molecular mechanisms, 2:488–489
neural systems, 2:484–485

Memory and knowledge, 1:144–145, 2:489–492
justification, 2:490–491
knowledge of the past, 2:491–492
roles of memory, 2:489–490
Memory codes, 2:766–767
Memory load, 2:790
Memory recall, dynamics, 2:492–496
See also Recall
associative asymmetry, 2:494
clustering, 2:495
contiguity, 2:493–494, 496
errors, 2:495
eyewitness memory, 1:340–342
interresponse times, 2:495–496
mnemonic strategies, 2:522–523
primacy, 2:493
recency, 2:492–493
retrieved context theory, 2:496
semantic proximity, 2:494–495

Memory spans, 1:14–15
Menary, Richard, 1:335, 338
Mendelian theory, 1:389
Mendl, Michael, 1:193
Menezes, Jean Charles de, 1:347
Mens rea, 1:95

Mental action, 2:497–498
Mental causation, 2:499–501
anomalous monism, 1:32–34

Mental effort, 2:502–504
central nervous system, 2:504
costs and allocation, 2:503
metacognition, 2:503–504
performance effects, 2:502
processing, 2:502–503
resources and mobilization, 2:502

Mental logic theory, 1:228
Mental models theory, 1:228–229
Mental representations
allocentric, 2:771
change blindness, 1:146–147
intension and extension, 1:426
intentionality, 1:140–143
mental transformations vs., 2:769
representational theory of mind, 2:652–654
spatial, 2:770–771
thinking, 2:740
visual imagery, 2:757–760

Mental rotation, 2:769
Mental state
functional vs. intrinsic properties of, 1:334
structure of, 2:513–514
Méré, Chevalier de, 1:220
Mere exposure effect, 2:754
Mere ownership effect, 1:322
Merikle, Philip, 2:751
Merker, Bjorn, 1:190
Message formulation, 2:616–617
Meta-analysis, 1:423, 2:505–507
Metacognition, 1:208, 2:503–504, 796
Metacognition and education, 2:508–509
Metacognitive knowledge, 2:508
Metacontrast, 2:760
Metaphor, 2:509–512
Metaphoric gestures, 1:372
Metaphysics, 2:554
Meter, 2:549, 618
Method of subtraction, 2:630–631
Methodological behaviorism, 1:110
Meyer, David, 1:62
Microcephalin, 1:371–372
Microcephaly, 1:371
Microgenetic theory, 2:513–515
Microgenetic theory, 2:513–514
Midbrain, 1:86–87
Migraine, 1:101
Milgram, Stanley, 2:588
Mill, John Stuart, 1:302, 2:526
Miller, E. K., 1:63
Miller, George, 1:14, 2:786
Miller, Jeff, 1:61–62
Miller, Kaarlo, 2:444
Millikan, Ruth, 1:38, 187, 2:703
Mills, Charles, 1:288
Milyavsky, Maxim, 2:778
Mind. See also Philosophy of mind; Theory of mind
Berkeley’s idealism, 1:397–398
electrophysiological studies, 1:273–277
extended mind, 1:335–338
intuitive vs. reflective, 2:741
mental action, 2:497–498
mental causation, 2:499–501
mental effort, 2:502–504
representational theory of, 2:652–654
teleology, 2:735–737
visual imagery, 2:757–760
Williams syndrome, 2:775–776
Mind mapping, 1:50
Mind-body problem, 2:516–520
anomalous monism, 1:32–34
behaviorism, 1:110
consciousness, 1:194–198
criticisms of, 2:447–448
current status, 2:518–520
dualism, 2:516–517
eliminativism, 2:517
folk theories, 2:517
functionalism, 2:517
knowledge-how, 2:447–448
mental causation, 2:499–501, 516–517, 519
personal identity, 2:582–585
physicalism, 2:517–519
Mind–brain identity theory, 2:638–639. See also Identity theory; Type identity theory
Mindfulness based cognitive therapy, 1:107
Minimal group paradigm, 1:433
Minimum alveolar concentration (MAC), 1:30
Minus span, 2:794
Mirror neurons, 2:520–521
Mirror systems, 1:163
Mischel, Walter, 2:587
Mismatch negativity, 2:698
Mismatching (gesture), 1:373
Mitchell, David, 2:752
Mithen, Steven, 2:548
Mnemonic strategies, 2:522–523
Mnemosyne project, 2:712
MODE (motivation and opportunity as determinants) model, 1:81
Modeling. See also Computational models
multinomial modeling, 2:538–541
natural action selection, 2:556–558
Modeling causal learning, 2:523–525
Modular-stores theory, 2:790
Molaison, Henry, 2:484–485
Moll, Jorge, 1:303
Monism, 1:190, 193, 305, 2:500
Monitoring, 2:508–509, 619
Monkeys, mirror neurons in, 2:520–521
Monozygotic twins reared apart (MZA), 1:390
Monozygotic twins reared together (MZT), 1:390
Montague, Richard, 1:424–425
Montaigne, Michel de, 2:589
Montrul, Silvina, 1:392
Mood disorders, 1:42
Moore, G. E., 1:151
Moral development, 2:525–527. See also Emotion and moral development
Moral judgment, 2:747
Moral sentiments, 1:376
Morality
action and responsibility, 2:597–599
attrition theory, 1:84
emotion, 1:300–304
guilt, 1:376–377
judgment vs. behavior, 1:300–301
Moran, Richard, 2:677
Moran, Seana, 2:543
Moratorium (identity), 2:586
Morgenstern, Oskar, 1:20, 222, 2:565
Morillo, Carolyn, 1:246
Morlaas, Joseph, 1:47
Morphemism, 2:618
Morphology, 2:458
Morris, Steve, 1:329
Morrison, Robert, 1:27, 28
Morton, Adam, 1:360
Morton, William, 1:30
Moscovitch, Morris, 1:262
Mother's sleep, 2:696
Motion. See Movement
Motion parallax, 1:234
Motivated thinking, 2:528–529
Motor learning, practical aspects, 2:529–532
Motor skill learning, 2:486
Motor system, development of, 2:532–535
creativity in, 2:534–535
malleability of, 2:535
prospective action, 2:533–534
ubiquity of movement, 2:533
Motor theories, 1:162
Movement
optic flow, 2:573–576
representation development, 2:655
ubiquity of, 2:533
Moving mask experiments, 1:339
Mozart, Wolfgang Amadeus, Così Fan Tutte, 1:148
Mu rhythm, 1:274
Müller, Georg, 2:481–482
Multidimensional scaling, 2:687–688, 690
Multimodal conversational interfaces, 2:536
Multimodal conversational systems, 2:535–538
interpretation function, 2:536–537
multimodal generation, 2:537
system architecture, 2:536
types, 2:536
Multinomial modeling, 2:538–541
case example, 2:539–540
common aspects, 2:540
current and future research, 2:540–541
data structure and model development, 2:538–539
Multinomial processing tree, 2:539
Multiple intelligences theory, 2:541–543.
See also Learning styles
Multiple trace theory, 1:24

See also Divided attention and memory

automation, 2:543

general principles, 2:546

neural mechanisms, 2:545

practice, 2:545

processing duration, 2:544–545

task switching, 2:544

threaded cognition, 2:545–546

types of processing, 2:544

Münsterberg, Hugo, 1:340, 2:482

Murdock, Ben, 2:495, 685

Murphy, Gregory, 1:171, 173, 180

Murray, Charles, 1:355

Music
defining, 2:549

definition, 2:549

emotion and, 2:550

evolution, 2:547–548, 550

Music and the evolution of language, 2:547–548

Music perception, 2:548–551

Müsseler, Jochen, 1:162

Mysterianism, 2:638

Mysticism, 2:628

Myth of the given, 2:489–490

Naccache, Lionel, 2:504

Nadelhoffer, Thomas, 1:329

Nagel, Thomas, 1:188, 286

Nahmias, Eddy, 1:329

Naïve realism, 2:553–554

Naltrexone, 1:10–11

Names, 1:424

Narcissistic personality disorder, 2:554–556

Nash equilibrium, 2:1223

National Academy of Sciences, 2:471, 472

National Epidemiologic Survey on Alcohol and Related Conditions, 1:9

Nativism

to knowledge acquisition, 2:449–450

language, 1:413–417, 2:455, 461

spatial cognition, 2:713

Natural action selection, modeling, 2:556–558

Natural assessments, 1:102, 2:637

Natural kind anti-individualism (NKA), 1:38

Natural language generation, 2:558–560


Nature vs. nurture, 1:414, 420–421

Navon, David, 1:61–62

N-choice reaction time tasks, 2:631–632

Neale, Stephen, 1:237

Negation parameter, 1:415–416

Negative affectivity, 1:42

Negative aftereffects, 2:480

Negative concord, 1:416

Negative health, 1:379

Negative heautoscopy, 1:101

Neisser, Ulric, 1:336, 404, 2:553–554, 753

Neumark, Lawrence, 2:448

Neo-empiricism, 1:174–175

Neural adaptation, in addiction, 1:10

Neural correlates of consciousness, 2:560–562

Neural networks, 1:129, 166

Neural reference space, 2:299–300

Neurodynamics of visual search, 2:562–564

Neuroeconomics, 2:564–567

Neuroimaging. See also Electrophysiological studies of mind

agging, 1:18–19

attention, 1:56–59

attention and emotion, 1:70–71

consciousness, 2:561

decision making, 1:217

dyslexia, 1:268–269

emotional recognition, 1:314, 316–318

exercise and the brain, 1:327–328

facial expression recognition, 1:353

inner speech, 1:418

intelligence, 1:420

memory, 1:402–403

mental effort, 2:504

neuroeconomics, 2:567

reasoning, 2:746–747

schizophrenia, 2:664–665

sleep, 2:696

Stroop effect, 2:723

synesthesia, 2:726

visual imagery, 2:758

visual working memory, 2:766–767

Neurological damage. See Brain damage

Neurology. See also Brain action selection, 2:557

attention and emotion, 1:70–71

divided attention and memory, 1:99–101

blind sight, 1:122–123

Capgras delusion, 1:127

categorization, 1:130–134

cognitive dissonance, 1:155

confabulation, 1:185

consciousness and embodiment, 1:196

decision making, 1:214–217, 248

desire, 1:245–246

dreams, 2:699

emotional recognition, 1:313–319

event memory, 1:325

explanatory gap, 1:333–335

face perception, 1:345–346

free will, 2:611–612

hearing, 1:85–88
human classification learning, 1:393–394
inhibition of return, 1:412–413
intelligence, 1:419–421
layered control architectures, 2:465–466
memory, 2:483–489
mental effort, 2:504
mirror neurons, 2:520–521
morality and emotion, 1:303
multiple intelligences, 2:542–543
multitasking, 2:545
optic flow, 2:575
reinforcement, 2:645–646
representational theory of mind, 2:653
schizophrenia, 2:663–666
scientific reasoning, 2:668
self-consciousness, 2:675
sleep, 2:697–698
stereopsis, 2:721–722
synesthesia, 2:726
taste, 2:732
time perception, 2:745
New theory of reference, 1:37
Newcomb's problem, 1:223
Newman, Matt, 1:210–211
Nichols, Shaun, 1:302, 328, 329, 358, 438
Nietzsche, Friedrich, 2:659
Nightmares, 2:699
Nisbett, Richard, 1:314, 135, 138
Nondeclarative memory, 1:24, 130
Nondirectional outcome-motivated thinking, 2:528–529
Nonhumans. See Animals
Nonlove, 2:473
Nonreductive physicalism, 2:639
Nonrelationalism, 2:703
Non-REM (NREM) sleep, 2:695–699
Nonverbal affect lexicons, 1:314
Norepinephrine, 2:504
Norman, Donald (Don), 1:6–7, 60, 258
Normativity, 2:599–600
Northoff, Georg, 1:191
Nosofsky, Robert, 1:173
Noun-noun combinations, 1:179–181
Novelty, 1:412
Nucci, Larry, 2:526
Numeric priming, 1:29
Nussbaum, Martha, 2:525, 526
Oaksford, Mike, 2:742
Object substitutions, 1:6, 8
Object-dependent thought, 1:240–241, 2:569–571
Objectivism, 2:701, 733–734
Objects, 2:515, 700–703
Objects of memory, 1:144–145, 2:571–572
O'Brien, David, 1:228
Observation probability density function, 2:478
Obsessive-compulsive disorder, 1:41, 2:558
Ochiba, Cynthia, 1:49
Oculomotor system, 1:412
Okada, Ron, 2:495
Oller, Kimbrough, 1:117
Olson, Eric, 2:583
Olson, Mancur, 1:160
Omission errors, 1:6, 7, 8
Omissions, 2:597–598
Operant conditioning, 2:644
Operant learning, 1:105, 110, 245
Operation span, 2:789, 794
Opioid receptor agonists, 1:10
Opioids, 1:307–308
Oppenheimer, Danny, 2:748
Optic flow, 2:573–576
Optimality, 2:557
Order memory. See Sequential memory, computational perspectives; Serial order memory, computational perspectives
Orthographic lexicon, 2:785
Orthonasal olfaction, 2:732
O'Shaughnessy, Brian, 2:498
Osman, Hussain, 1:347
Osman, Magda, 2:748
Ossicles, 1:85, 383
Ostension, 2:780
O'Sullivan, Maureen, 1:121
Other-race identifications, 1:341, 342, 348
Otto thought experiment, 1:336–338
Outcome conflict, 1:61
Outcome value, 2:709
Out-of-body experiences, 1:100–101
Overdetermination, 2:500, 517, 598, 602
Overlearning, 1:242
Overregulation in language development, 2:460–461
Oviatt, Sharon, 2:536
Ozbay, Erkut, 2:567
Pain, 1:427–428
Pair-clustering model, 2:539–540
Paired-associate learning, 1:393
Panic disorder, 1:40
Panini, 2:682
Panksepp, Jaak, 1:190–192, 306
Panpsychism, 1:399
Panum's fusional area, 2:720
Papez, James, 1:314
Papineau, David, 1:38
Paracontrast, 2:760
Paradigmatic associations, 2:683
Paradoxical sleep. See Rapid eye movement (REM) sleep
Parallel distributed processing, 1:279
Parallel processing, 2:544, 562, 616, 624, 722, 765
Parameters, language development and, 1:413–417
Paré, Denis, 2:698
Parfit, Derek, 2:583
Parietal cortex, 1:413
Parity principle, 1:336
Parkin, Alan, 1:16
Parkinson’s disease, 1:132, 231, 246, 2:486, 487, 558
Parks, Emily, 1:58
Parrot, Gerrod, 1:323
Pascal, Blaise, 1:151, 220–221
Pascal’s Wager, 1:222
Pashler, Harold (Hal), 1:61, 261, 262, 263, 2:623, 777
Passingham, Richard, 1:123, 2:772, 773
Past verb tense, 2:460
Pastoral counselors, 2:651–652
Path analysis, 1:389
Path perception, 2:574–575
Pattern masking, 2:754, 760
Pavlov, Ivan, 1:105, 2:605, 643–644, 687
Pavlovian conditioning, 2:487
Payne, John, 2:503
Peacocke, Christopher, 1:112, 151, 172, 230, 2:498, 677
Pearson, Karl, 2:506
Peer relationships, 2:649
Peg words, 2:522
Peirce, Charles, 2:513
Peirce, Charles Sanders, 2:724, 752
Pennebaker, James, 1:210
Perceptual consciousness and attention, 2:577–580.
See also Unconscious perception
See also Visuospatial reasoning
size constancy, 2:580–582
Perceptual learning, 1:250–251, 2:717–718
Perceptual loop hypothesis, 2:619
Perceptual memory, 2:767
Perceptual priming, 2:486
Perceptual representation, 1:174, 2:654–656
Perceptual representation memory system, 1:133, 402
Perceptual span, 1:339
Performance
mental effort and, 2:502
multitasking and, 2:543–546
reaction time, 2:630–633
rehearsal and memory, 2:641–643
skill learning, 2:691–694
Performance expectancy, 2:709
Performance operating characteristics (POCs), 1:60–61
Perky effect, 2:773
Permastore, 1:16
Perry, John, 1:407
Perseverative errors, 1:6, 8
Person memory, 2:704–706
Personal equation, 2:630
Personal identity, 2:582–585. See also
Consciousness and embodiment
animalism, 2:584–585
concept of identity, 2:582–583
concept of person, 2:583–584
dualism, 2:583
See also Self, development of
Personal intentions accounts of joint intention, 2:444–445
Personal semantics, 1:24
Personality. See also Character and personality, philosophical perspectives
attraction, 1:82–83
consistency/stability of, 2:587–588
ideology and, 2:609
mental effort, 2:503
music, 2:550–551
Personality disorders, 2:555
Personality: individual versus situation debate, 2:587–588. See also Attitudes and behavior;
Character and personality, philosophical perspectives
attributions, 2:587
consistency of personality, 2:587–588
interactionism, 2:588
Perspective, 1:232–233
Perspective size hypothesis, 2:582
Perspective taking in language processing, 2:588–590
Persuasion, 2:590–594. See also Attitude change
attitude change, 1:78
central vs. peripheral routes, 2:591–592
compliance vs., 2:591
dual process models, 2:591–592
factors influencing, 2:592
overview, 2:590
Pervasive developmental disorders (PDDs), 1:90
Pesendorfer, Wolfgang, 2:567
Pessimistic incompatibilists, 1:361, 363
Pettigrew, John, 2:721
Petty, Richard, 2:591
Phantom limb sensations, 1:427
Pharmacotherapy
addiction, 1:10–11
anxiety disorders, 1:42
autism, 1:93
delusions, 1:232
psychopathologies, 1:305
schizophrenia, 2:666
Phase locking, 1:384
Phelps, Elizabeth, 2:567
Phenomenal awareness, 2:677
Phenomenal concepts, 1:286–287, 2:603
Phenomenal consciousness, 1:1–3, 2:561–562, 602
Phenomenal properties, 1:286
Phenomenalism, 1:398–399, 2:637
Phenomenology
emotions, 1:431
experience, 2:635
hearing, 1:388
olfactory experience, 2:700–701
theory of appearing, 2:737–739
thinking, 1:188
Phenomenology of action, 2:595–596
Phenomenon of the missing fundamental, 1:385
Phenotype, 2:775
Philosophy
character and personality, 1:147–150
concepts, 1:52–53, 150–152, 166–168, 171–175
decision theory, 1:220–224
experimental philosophy, 1:328–330
hearing, 1:387–389
know-how, 2:447–448
naïve realism, 2:553–554
reduction, 2:636–637
smell, 2:700–704
taste, 2:731–734
Philosophy of action, 2:596–600
moral responsibility, 2:597–599
reasons for action, 2:599–600
revival of, 2:597
teleology, 2:735–737
Philosophy of language, 1:424–426
Philosophy of mind, 1:426, 2:443, 448, 636–639, 735–737
Philosophy of perception, 2:553–554
Phlogiston, 1:278
Phonemes, 1:267–268, 2:618, 717, 784
Phonemic restoration, 2:718
Phonetic variation, 2:717
Phonological development, 2:456–457
Phonological dyslexia, 1:266
Phonological encoding, 2:619
Phonological loop, 2:767, 791–792
Phonological processing, 2:269–271
Phonological representation, 1:267–268
Phonological similarity effect, 2:685
Phonological structure, 2:464
Physicalism, 2:600–603, 637. See also Materialism
argument for, 2:602
criticisms of, 2:448
emergent, 2:601–602
explanatory gap, 2:602–603
mental causation, 2:500–501
mind-body problem, 2:517–519
nature of, 2:600–602
olfactory experience, 2:703
problems for, 2:602–603
token-token physicalism, 2:518
Physics
children's theories of, 2:450–451
physicalism and, 2:600–601
Picard, Rosalind, 1:174
Pick, Arnold, 1:47, 183
Pick's disease, 1:48
Piddins, 1:118–120
Piéron's law, 2:632
Pierrehumbert, Janet, 1:268
Pilzecker, Alfons, 1:10–11
Place, U. T., 1:189
Place learning, 2:714, 715
Place of articulation, 2:717
Placebo effect, 2:603–606
background, 2:603–604
factors in, 2:604–605
individual differences, 2:605
theories, 2:605–606
Placebo surgery, 2:604–605
Planned behavior theory, 1:81
Planning, two-system model of, 2:747
Planning in language production, 2:607–608
Plato, 1:147, 151, 2:499, 513, 608
Platt, Michael, 2:566
Plaza, Enric, 1:129
Pleasure, 1:246, 380
Plotinus, 2:580
Plural nouns, 2:460–461
Poe, Edgar Allan, 1:102
Poeck, Klaus, 1:47
Poldrack, Russell, 2:567
Polinsky, Maria, 1:392
Poliopic heautoscopy, 1:101
Political psychology, 2:608–611
Polygraphy, 2:470–472
Polyn, Sean, 2:566
Polysomnographic recordings, 2:696
Pop-out search, 2:562
Popper, Karl, 2:667
Population coding, 1:279
Porter, Bruce, 1:128
Positional tagging models, 2:686
Positive aftereffects, 2:480
Positive health, 1:379
Positron emission tomography (PET), 1:57–58, 70, 273, 316, 402, 420, 2:696, 726, 745
Posner, Michael I., 1:410, 411
Possible worlds, 1:425
Post completion errors, 1:6
Post concussive syndrome, 1:125
Posterior parietal lobe, 1:183
Postevent information, 1:340–341
Post-traumatic stress disorder (PTSD), 1:41, 125
Pound, Roscoe, 2:469
Pourtois, Gilles, 1:317
Power, Jonathan, 2:543
Power law of forgetting, 2:692
Power law of practice, 2:692
Power PC theory, 2:524
Practical wisdom, 1:149
Practice
distribution of, 2:693
mental, 2:693–694
motor learning, 2:530
multitasking, 2:545
skill learning, 2:692–693
spacing effect, 2:710–713
variability of, 2:692–693
Pragma, 2:473
Pragmatics, in language development, 2:456
Preconscious free will, 2:611–612
Predicates, 1:424
Prediction, 2:633
Prediction, clinical versus actuarial, 2:613–615
Predictive coding, 1:64
Predictive inferences, 1:409
Prefrontal cortex, 1:15, 63, 130–132, 184, 325, 2:723
Prehension, prospective control of, 2:533
Prelec, Drazen, 2:566
Premack, David, 1:168
Premotor cortex damage, 1:46
Presence, feeling of, 1:101
Presentism, 1:399
Preston, Tim, 2:721
Prevention-focused thinking, 2:528
Primacy effect, 2:493, 685, 788–789
Primacy gradient models, 2:685–686
Primary progressive aphasia, 1:45–46
Priming, 1:31, 2:577, 579–580, 706, 707
Principal component analysis (PCA), 1:347–348, 350
Principle of charity, 1:33
Principles and parameters theory, 1:413–414
Pringle-Morgan, William, 1:269
Prinz, Jesse, 1:66, 174–175, 431, 2:629
Prinz, Wolfgang, 1:161, 162, 2:772
Prior, Anat, 1:116
Prior knowledge, 2:524
Prisoner's Dilemma, 1:159–160, 223, 436
Proactive interference, 2:480–483
Pro-attitudes, 1:331–332
Probability
availability heuristic, 1:102–103
decision making, 1:221
degrees of belief, 1:264–265
representativeness heuristic, 2:657–658
thinking, 2:742
Problem categories, 1:137
Problem solving, 2:747, 759–760
Problem distribution regularity, 1:129
Problem-solution regularity, 1:129
Procedural information, 2:691–692
Procedural memory, 1:130, 132, 224–226, 402
Procedural reinstatement, 2:692
Processing duration, 2:544–545
Processing methods, 2:544
Processing resources, 1:60–62, 2:544
Production of language, 2:615–619. See also entries
beginning Language production
agreement, 2:461–463
function assignment and constituent assembly, 2:618–619
gender differences, 1:369–370
incremental processing, 2:463–464
lexical selection and retrieval, 2:617–618
message formulation, 2:616–617
monitoring, 2:619
natural language generation, 2:558–560
overview, 2:615–616
phonological encoding, 2:619
Prosody in production
Prosodic boundaries, 2:124
Prosody, 2:703
Prosodic structure, 2:528
Proper linear models, 2:613–614
Properties
perception by smell, 2:703
perception by taste, 2:731
Property interpretation, 1:179–180
Proponent effect, 1:257
Propositional appeals, 1:78
Propositional awareness, 2:677–678
Propositional memories, 2:571–572
Propositions, 1:111
Propositions, as semantic knowledge, 2:680–683
Proprioception, 2:596
Prosodic boundaries, 2:620–621
Prosodic structure, 2:620–621
Prosody, 1:318, 2:549
Prosody in production, 2:619, 620–621
Prosopagnosia, 1:2, 3, 127, 346–347, 2:752
Prospect theory, 1:21
Prospective timing models, 2:744–745
Protolanguage, 2:547–548
Prototypes, 1:135, 138–139, 173, 2:474
Psychiatry, religion and, 2:650–652
Psychogenic illness, 2:605
Psycholinguistics, 1:44–45
Psychological behaviorism, 2:517
Psychological blows, 1:94
Psychological essentialism, 1:171
Psychological refractory period, 2:622–624
bottleneck models, 1:62, 65, 2:623–624
capacity-sharing models, 2:622–623
exceptions to, 2:623–624
multitasking, 2:544–545
Psychology
children's theories of, 2:451
neuroeconomics, 2:564–567
political psychology, 2:608–611
Psychopathology
emotion, 1:303–308
morality, 1:303–304
narcissistic personality disorder, 2:555
Psychophysical scaling, 2:751
Psychophysical tuning curve, 1:89
Psychosocial interventions, 2:666
Psychotherapy, 1:10
Proust, 2:580
Public goods, 1:160
Pulvinar, 1:123
Pure alexia, 1:44, 266
Pure dissent effect, 1:257
Pure motor speech disorders, 1:44
Pure word deafness, 1:44
Purkinje cells, 2:487
Putnam, Hilary, 1:37, 38, 52, 151–152, 178, 187
Pylyshyn, Zenon, 2:758
Qualia, 2:519, 635–636
Quality of life, 1:379
Quine, W. V., 1:151
R. v. Parks, 1:94
R. v. Quick, 1:95
R. v. Rabey, 1:94
R. v. Tolson, 1:94
Rabbitt, Patrick, 1:18
Rachlin, Howard, 1:110
Radical behaviorism, 1:108, 110
Ramachandran, Vilayanur, 2:726
Ramsey, Frank, 1:20, 222
Randomized controlled trials, 2:603–604
Randomized models of prediction, 2:614
Rapid automatized naming, 1:268
Rapid eye movement (REM) sleep, 2:695–699
Rapid serial visual presentation, 1:74
Ratcliff, Roger, 1:220
Rational actor hypothesis
Allais paradox, 1:19–21
degrees of belief, 1:264–265
Dictator Game, 1:247–249
Rational choice theory, 1:220–221
Rationality. See also Reasoning; Thinking
attribution theory, 1:83–84
constraints on, 2:627–628
decision making, 1:212
ecological, 2:557
expectation and, 2:629
intentionality and sensation, 2:627
memory and, 2:629
nature of, 1:187
teleology, 2:736
theories of, 1:186–187
Rationality of emotion, 2:627–630
Raven's Progressive Matrices (RPM), 1:356, 421–423
Rawls, John, 2:526
Realistic conflict theory,
Recognition
Realism and instrumentalism
Reading span,
Reciprocity, persuasion based on,
Reaction time, 2:630–633
agaging, 1:17–18
history, 2:630–631
influences on, 2:632–633
types of experiments, 2:631–632
Read, Don, 2:753
Readiness potential, 1:277, 2:612
Reading
automaticity, 1:96, 2:722
dyslexia, 1:266–271
eye movements, 1:338–340
word learning, 2:782
word recognition, 2:785–786
Reading comprehension, 1:267
Reading span, 2:794
Realism
legal realism, 2:469
naive realism, 2:553–554
olfactory experience, 2:703
Realism and instrumentalism, 2:633–636
overview, 2:633–634
propositional attitudes, 2:634–635
qualia, 2:635–636
Realistic conflict theory, 1:434, 435
Realizer functionalism, 2:518–519
Reason, James, 1:6
Reasoned action theory, 1:80–81
Reasoning. See also Rationality; Scientific reasoning; Thinking
analogueal, 1:25–26, 2:468
argument mapping, 1:50–52
case-based, 1:128–130
deductive, 1:226–230, 2:468–469, 742
deliberative, 2:745–748
general- and special-purpose mechanisms, 1:229
inductive, 1:226
intuitive, 2:745–748
legal, 2:468–470
two-system models, 2:745–748
visuospatial, 2:769–772
Reasons
explanation of action, 1:331
normative vs. operative, 1:331
Reasons for acting, 2:599–600
Reboul-Lachaux, Jean, 1:127
Recalitrant emotions, 2:629
Recalibration, 2:717–718
Recall, 1:16. See also Memory recall, dynamics
Recency effect, 2:492–493, 685, 789
Receptive vocabulary, 2:779
Reciprocity, persuasion based on, 2:594
Recognition
Capgras delusion, 1:127–128
emotional, 1:313–319
face recognition, 1:343–344
memory, 1:16
Recollection, 1:24
Reduction, 2:636–637
Reductive physicalism, 2:636–639. See also Physicalism
behaviorism, 2:638
evaluation of theories, 2:638–639
mind-brain identity theory, 2:638
mysterianism, 2:638
reduction, 2:636–637
Referential communication, 1:203
Reflective mind, 2:741
Refractory period. See Psychological refractory period
Regret, 1:21, 2:639–641
Rehearsal and memory, 2:641–643. See also Retrieval practice (testing) effect; Spacing effect
Reinforcement, 1:9–10. See also Rewards
Reinforcement learning
brain, 1:132
decision making, 1:218–219
Reinforcement learning, psychological perspectives, 2:643–646
Reinforcement sensitivity theory, 1:315
Reinforcers, 1:109, 2:644, 645
Relation interpretation, 1:179–180
Relational similarity, 1:25, 27
Relationalism, 2:703
Relationships, development of, 2:646–650. See also Interpersonal relationships
attachment, 2:647–648
attraction, 1:82–83
friendships, 2:648–649
internal working models, 2:648
overview, 2:646–647
peer relationships, 2:649
romantic relationships, 2:649–650
Relative deprivation, 2:659
Relialibilism, 1:229
Reliability (research), 2:467
Religion, consciousness and embodiment, 1:194–195
Religion and psychiatry, 2:650–652
REM sleep. See Rapid eye movement (REM) sleep
REM sleep-on cells, 2:695
Reminiscing style, 1:326
Rensink, Ronald, 2:577
Repetition, mnemonic benefits of, 2:661
Replication, 2:506–507
Representation. See also Mental representations
analogueal, 1:174
conceptual, 2:654–656
facial, 1:347–348
perception as, 1:255–256
perceptual, 1:174, 2:654–656
phonological, 1:267–268
smell and, 2:700–701
Representation theorems, 1:221–222
Representation theorems
Chinese room argument, 2:653–654
connectionism, 2:653
historical versions, 2:652
intension and extension, 1:426
recent versions, 2:652–653
Representationalism, 1:188–189
Representations, development of, 2:654–656.
See also Concepts, development of
Representativeness heuristic, 2:657–658
Rescorla, Robert, 2:643
Rescorla-Wagner model, 2:524
Research results, 2:505–506
Resentment, 2:658–660. See also Envy
Resistance to change, 1:153
Resonance, 2:716
Resource theories of attention, 1:60–62, 75
Response expectancy, 2:605–606
Response learning, 2:714
Ressentiment, 2:659
Retentive reduction, 2:638
Retrieval practice (testing) effect, 2:660–661. See also Rehearsal and memory; Spacing effect
Retrieved context theory, 2:496
Retroactive interference, 2:481–482
Retrograde amnesia, 1:24–25, 95, 2:485
Retronasal olfaction, 2:732–733
Rett's disorder, 1:90
Reward learning, 1:245
Reward prediction error, 1:215
Rewards. See also Reinforcement
decision making, 1:215–216, 218–220
desire, 1:245–246
Rey, Georges, 1:171
Rhymes, as mnemonic device, 2:523
Rhythm, 2:549
Ribot, Théodule-Armand, 2:481
Richardson, Daniel, 1:205
Richardson, Lewis F., 1:435
Richardson, Lucy, 2:690
Riddoch, George, 1:121
Riefer, David, 2:539
Right hemisphere of brain
apraxia, 1:46–49
attention, 1:58
confabulation, 1:185
emotion, 1:287–289, 314–315
language processing, 2:456
Right-ear advantage, 2:718
Rights, 2:527
Right-wing authoritarianism (RWA) scale, 2:610
Rips, Lance, 1:228, 2:688
Risk shift, 1:374
Rissland, Edwina, 1:128
Rizzolatti, Giacomo, 1:163
Roberts, Robert, 1:432
Robinson, Terry, 1:10
Robotics, 1:196, 2:465, 536, 668
Rock, Irvin, 1:404, 2:577
Rodin, Judith, 1:324
Rohrer, Doug, 1:262
Role functionalism, 2:518–519
Rolls, Edmund, 2:563
Romantic love, 2:474
Rosch, Eleanor, 1:167, 173, 2:474
Rosenberg, Morris, 2:567
Rosenthal, David, 1:3
Rosenthal, Robert, 1:354
Ross, Elliott, 1:315
Rotational flow, 2:573–574
Rothi, Leslie, 1:49
Rotor-pursuit task, 2:486
Rousseau, Jean-Jacques, 2:608–609
Routine system, 1:7
Rowlands, Mark, 1:335, 338
Rozin, Paul, 1:301
Rubin, David, 1:16
Rule-based category learning, 1:132
Rumelhart, David, 2:646, 785
Rupert, Robert, 1:142
Rushton, Simon, 2:575
Russell, Bertrand, 1:151, 236–239, 241, 2:452–453, 570–571
Russell, James, 1:352
Rutter, Michael, 1:92
Ryle, Gilbert, 2:447–448
Saccades, 1:338–339
Sacks, Harvey, 1:203
Salmoni, Alan, 2:531
Salovey, Peter, 1:324
Salthouse, Tim, 1:14, 18
Samuelson, Paul, 2:565
Sandwich masking, 2:754
Sane automatism, 1:95
Satisfaction, 1:380–381
Saults, J. Scott, 2:790
Savage, Leonard J., 1:20, 222, 2:565
Savin, N. E., 1:247
Scalar expectancy theory, 2:744
Scarcity, persuasion based on, 2:593
Schaeken, Walter, 1:227
Schank, Roger, 1:128
Schegloff, Emanuel, 1:203
Scheines, Richard, 1:329
Scheler, Max, 2:639
Schemas, 2:704
Schiffer, Stephen, 1:241
Schizophrenia, 2:663–666
  action selection, 2:558
  aphasia, 1:43
  autoscopic phenomena, 1:101
  delusions, 1:231
  neurodevelopmental model, 2:663–664
  onset of, 2:664–665
  pathology, 2:665–666
  risk factors, 2:664–665
  symptomology, 2:663
  treatment, 2:666
Schmid, Michael, 1:123
Schmidt, Richard, 2:531
Schnall, Simone, 1:303
Schoenfeld, Alan, 2:508
Schotter, Andrew, 2:567
Schroyens, Walter, 1:227
Schultz, Wolfram, 1:245, 2:645
Schwarz, Norbert, 2:504
Schwitzgebel, Eric, 1:328
Scientific reasoning, 2:667–668
  analysis of, 2:667–668
  children's, 2:450–451
  historical context, 2:668
  legal reasoning vs., 2:469–470
  modern approaches, 2:668
  overview, 2:667
  role of hypotheses, 2:667, 668
Scientism, 1:144
Scotomas, 1:120–121
Scott Morton, Michael S., 1:213–214
Scottish Enlightenment, 1:301
Seamon, John, 2:752
Searle, John, 1:2, 188, 200–201, 2:444, 510, 570, 572, 595, 653–654
Sears, David, 2:610
Schantz, Natalie, 1:162
Sebat, Jonathan, 1:92
Secondary qualities, 2:703, 734
Secure attachment style, 2:475, 647
Seton, Martin, 1:247
Segal, Gabriel, 1:238
Segmentation (speech), 2:717–718
Seidenberg, Mark, 1:268
Selective attention, 1:53–56, 75, 411
Self. See also Personal identity
  autoscopic phenomena, 1:99–102
  cognitive dissonance, 1:154–155
  indexical thought, 1:406–407
  origin of, 2:515
Self, development of, 2:669–673. See also Moral development; Personal identity, development of
  Self-actualization, 1:380
  Self-affects, 2:670
  Self-affirmation theory, 1:155
  Self-awareness, 2:674
  Self-categorization, 1:435
  Self-cognitions, 2:707
Self-consciousness, 2:673–675
  guilt, 1:376
  meanings, 2:674
  neural bases, 2:675
Self-efficacy, 2:509
Self-esteem, 2:669, 671–673
Self-help audio media, 2:753, 754
Self-knowledge, 2:676–678
  belief and judgment, 1:112
  constitutive accounts, 2:677
  epistemic accounts, 2:677
  inferential model, 2:677
  introspection, 1:436–439
  introspective model, 2:676–677
  mental action, 2:498
  wakeful consciousness, 2:498
Self-monitoring, 1:438
Self-motion, 2:573–576
Self-perception theory, 1:78, 437
Self-regulation, 1:106, 310, 2:502
Sellars, Wilfrid, 1:358, 359, 2:444
Semantic dementia, 1:45, 2:678–679
Semantic externalism, 1:152
Semantic knowledge, 2:680–683
Semantic memory, 2:679–681
  aphasia, 1:45
  defined, 1:402, 2:679
  feature overlap model, 2:680–681
  hierarchical model, 2:680
  semantic dementia, 2:678
  spreading activation model, 2:681
  time frame, 1:24
Semantic memory, computational perspectives, 2:681–683
Semantic priming effects, 1:45, 2:618
Semantic proximity, 2:494–495
Semi-Automatic Ground Environment (SAGE) system, 1:213
Sensation
  awareness of, 2:752
  cross-modal effects, 2:732–734
Sense data, 1:398, 2:452
Sensibility theory, 1:302
Sensory adaptation, 2:478–480
Sensory memory, 2:788
Sensory-motor integration, 1:183
Sensory-processing theories, 1:268
Sentence onset latency, 2:464
Sentimentalism, 1:301–302
September 11, 2001 attacks, 1:16
Sequential memory, computational perspectives, 2:683–685
Sequential models of causal learning, 2:524–525
Sequential probability ratio test, 1:219–220
Sequential sampling models, 2:633
Serial order memory, computational perspectives, 2:685–686
Serial processing, 2:544–545, 562, 623–624, 765
Serial reaction time task, 2:486
Serial search, 2:562
Sex, gender vs., 1:369
Sexual system, and love, 2:475
Shallice, Tim, 1:7
Sham surgery, 2:604–605
Shame, 1:376
Shape, 1:234–235
Shape constancy, 2:580
Shared environmental variance, 1:389–390
Sharma, Dinkar, 1:73
Sharvy, Richard, 1:239
Shaver, Phillip, 2:475
Sidanius, James, 1:435
Shizgal, Peter, 2:566
Shoben, Edward, 1:180, 2:688
Shoemaker, Sydney, 1:438, 2:677
Short-term memory, 1:182, 2:766, 787–788
Sidman, James, 2:610
Siegel, Linda, 2:796
Siewert, Charles, 1:188
Sighted blindness, 2:578, 579
Signal and response repetition effect, 1:17
Similarity, 2:687–691. See also Analogical mapping and reasoning; Analogical reasoning, models of development
attraction, 1:83
classification, 1:135–136
concept of, 2:690–691
featural, 1:25–26
models, 2:687–690
overview, 2:687
relational, 1:25, 27
Simmel, Georg, 1:434
Simon, Herbert, 1:212, 213, 439
Simon, Théodore, 1:422
Simonides, 2:522
Simons, Daniel, 2:577
Simple feelings, 1:299
Simple recurrent networks, 1:8
Simulation theory, 1:359
Simultaneous contradictory belief, 2:747
Simultaneous masking, 2:760
Singer, W., 1:64
Single fixation duration, 1:339
Single photon emission computerized tomography (SPECT), 1:316
Single-bottleneck model, 1:61–62, 65
Singular thought, 1:240–242, 2:569–571
Singular thought anti-individualism (STA), 1:38–39
Singular value decomposition, 2:688
Situated action, 1:258
Situated cognition, 1:258
Situated dialogue agents, 2:536, 537
Situationism, 1:147–149
Size constancy, 2:580–582
Size-distance invariance hypothesis, 2:581
Skaggs, Ernest Burton, 2:481
Skeletal muscle, 2:773
Skill learning, 2:486
Skill learning, enhancement of, 2:691–694
Skin conductance response, 2:295
Skinner, Erin, 1:263
Skyrms, Brian, 1:265
Social categories, 1:137
Slant constancy, 2:580
Sleep and dreams, 2:694–699
dream mentation, 2:698–699
information processing, 2:696–698
learning and, 1:251–252
methods of studying, 2:695–696
overview, 2:694–695
Sleep spindles, 2:697
Slow wave activity, 1:275
Smallman, Harvey, 2:582
Smart, J.J.C., 2:600
Smell, philosophical perspectives, 2:700–704, 732
Smetana, Judith, 2:526
Smiling, 1:282
Smith, Adam, 1:376
Smith, Andrew, 2:575
Smith, Ed, 2:688
Smith, Michael, 1:244
Smith, Richard, 1:323
Smith, Susan, 1:210
Smith, Vernon, 2:566
Smooth muscle, 2:773
Snedeker, Jesse, 2:621
Snowdon, Paul, 2:583
Social anti-individualism (SA), 1:39
Social anxiety disorder, 1:40
Social categories, 1:137
Social choice theory, 1:224
Social cognition, 2:704–708
category accessibility, 2:706
disciplinary content of, 2:706–707
dual-process theories, 2:706–707
mirror neurons, 2:520–521
origins, 2:704–706
other approaches, 2:707–708
person memory, 2:704–706
social phobia, 2:665
Social comparison, 1:374–375, 2:670
Social competence, 2:520–521
Social identity, 1:435
Social influence, 2:590–591
Social intuitionism, 1:302
Social learning, 2:451, 713
Social loafing, 2:708–710. See also Free riding
Social phobia, 1:40
Social psychology, 2:706
Social roles, 2:586
Social skills training, 2:666
Social validation, persuasion based on, 2:593–594
Social-cognitive theory, 2:442
Society for Neuroeconomics, 2:567
Socratic dialogue, 1:106
Somatic marker hypothesis, 1:296–297
Sommer, Marc, 1:56
Somnambulism, 1:94
Soteroiu, Matthew, 1:112
Soul, 2:583
Sound, 1:383–387, 2:548–549
Sound intensity, 1:88–89
Sound localization, 2:457
Source analog, 1:25
Source memory, 1:16
Source-filter theory, 2:716
Source-monitoring experiment, 2:538
Space. See also entries beginning with Spatial
descriptions of, 2:771
emotions associated with, 2:771–772
graphic representation of, 2:771
hearing, 1:387–388
mind/brain and, 2:515
representation development, 2:655–656
Spacing effect, 2:661, 710–711
Spacing effect, practical applications, 2:712–713
Spared priming, 1:31
Sparse nonnegative matrix factorization, 2:682
Spatial cognitio, development of, 2:713–716.
See also Visuospatial reasoning
developmental changes, 2:715–716
spatial coding system, 2:714–715
theoretical issues, 2:713–714
Spatial imagery, 2:759
Spatial mental representations, 2:770–771
Spatial neglect, 1:413
Spearman, Charles, 1:421–423
Specific language impairment (SLI), 2:459
Specific phobia, 1:41
Spectral edge frequency, 1:31
Speech
gestures and, 1:373
hearing and, 1:388–389
inner, 1:418–419
machine speech recognition, 2:477–478
Speech comprehension, 1:156–158, 252–253
aids to, 2:717–718
complicating factors, 2:717
machine speech recognition, 2:477–478
notable phenomena, 2:718
speech signal, 2:716–717
word recognition, 2:783–784
Speed-accuracy trade-off, 1:220, 2:694, 763
Sperber, Dan, 2:511
Spinoza, Baruch, 1:151, 194, 2:629
Spiral ganglion cell, 1:86
Splay angle, 2:576
Splitting, 1:124
Spoken dialogue systems, 1:204
Spreading activation model of semantic memory, 2:681
Spreading of alternatives, 1:154
Srull, Thomas K., 2:705
St. Petersburg Paradox, 1:222–223
Stampe, Dennis, 1:143
Stanford-Binet IQ test, 1:355–356, 422
Stevenson, Adlai, 1:213
Stanovich, Keith, 2:741, 796
Startle, 1:296
State transition probability matrix, 2:478
Statistical prediction rules, 2:613–615
Stavy, Ruth, 2:748
Steele, Claude, 1:155
Stein, John, 1:268
Stereacuity, 2:721
Stereopsis, 1:234, 2:719–722. See also Depth perception
Stereoscopes, 2:720
Stereotyping, 1:170, 207
Stereotypy, 1:109
Stern, Josef, 2:510
Sternberg, Saul, 2:631, 632
Stevenson, Adlai, 1:213
Stewart, Erika, 1:115
Steyvers, Mark, 2:682
Stich, Stephen (Steve), 1:329, 358, 438
Stimulus onset asynchrony (SOA), 2:760, 762
Stimulus-organism-response (S-O-R) chains, 1:106
Stimulus-response (S-R) chains, 1:105, 109–110, 111
Stimulus-response compatibility, 2:632
Stoeg, Petra, 1:121, 2:579
Stoel, Gjibert, 1:162
Stoicism, 2:629
Stone, Leland, 2:574
Stop-signal tasks, 2:632
Storage, compound word processing and, 1:163–165
Storge, 2:473
Strait, Dana, 2:542
Strange situation, 2:647–648
Strategic theories, of spacing effect, 2:711
Strategy-motivated thinking, 2:528
Strawson, Galen, 1:188, 399
Strawson, Peter, 1:237, 241, 244, 388
Stress-inoculation training, 1:106
Striate cortex, 1:57–58
Striatum, 1:131–133, 393, 2:486
Stroke, 1:181
Stroop, John Ridley, 2:722
Stroop effect, 2:722–723, 726
Stroop task, 1:18, 69, 70, 72–73, 96–97
Structural priming, 2:618
Structural theories, of knowledge acquisition, 2:450
Structure-mapping engine (SME), 1:28
Subcortical structures, 1:412–413
Subjective doubles, 1:364. See also Doppelgängers
Subjective expected utility, 1:221
Subjective threshold, 2:751
Subjective we-intentions, 2:444
Subjective well-being, 1:380–381
Subjectivism, 2:701, 733–734
Subjects, 2:514–515
Subliminal influence, 1:78, 2:750, 754
Subliminal perception. See also Unconscious perception
Subliminal priming, 2:707
Subset principle, 1:414, 417
Substance, 1:194
Substance abuse, 1:8–9
Substance dependence, 1:9
Substance use disorders (SUDs), 1:8–10
Substitution therapies, 1:10
Subsumption architecture, 2:465
Subtheory of love as a story, 2:474
Sucker effect, 2:709
Sumner, William Graham, 1:435
Superior colliculus, 1:123
SuperMemo, 2:712
SuperTasters, 2:733
Supervenience, 1:33, 2:501, 601
Suppressibility, of voluntary actions, 2:772
Surface dyslexia, 1:266
Surowiecki, James, 2:777
Sutton, Richard, 1:218, 338
Swain, Stacey, 1:330
Syllables, 2:618
Symbol learning, 2:779
Sympathy, 1:301
Synesthesia, 2:725–727
Syntactic production, agreement in, 2:618–619, 727–729. See also Language production, agreement in
Syntagmatic associations, 2:683
Syntax, 2:458
Sytsma, Justin, 1:329
Szabó, Zoltán, 1:237, 238
Tabert, Matthias, 1:318
Tachistoscopic viewing task, 1:315
Tallal, Paula, 1:268
Tan, Lydia, 2:642
Tanenhaus, Michael, 1:205, 367
Tangney, June, 1:376
Target analog, 1:25
Task switching, 2:544, 632
Task-positive network, 2:504
Taste, philosophical perspectives, 2:731–734
Tau rhythm, 1:274
Taylor, John, 1:73
Taylor, Stephen F., 1:318
TEACCH, 1:93
Teleology, 2:735–737
causalist view vs., 2:735–737
evolution, 2:737
explanation of action, 1:332–333
reductive accounts, 2:736–737
Teleosemantics, 1:143
Telford, Charles, 2:622
Template matching, 2:477–478
Tempo, 2:549
Temporal cortex, 2:485
Temporal discounting, 1:216
Temporal lobe, 1:325
Temporally graded retrograde amnesia, 2:482
Tenenbaum, Joshua, 2:682
Tesan, Graciela, 1:415
Testimony, eyewitness, 1:340–342
Testing effect, 2:660–661
Tetlock, Philip, 2:610
Text learning, 2:795–796
Thalamus, 1:22, 87, 2:695, 697
Thematic Apperception Test, 2:609
Theoretical behaviorism, 1:108
Theories, 1:173
Theory construction, in child development, 1:170
Theory of appearing, 2:737–739
Theory of dimensional overlap, 1:162
Theory of mind
children, 2:451
introspection, 1:438
nonhumans and, 1:168
schizophrenia, 2:665
Theory of planned behavior, 1:81
Theory of reasoned action, 1:80–81
Theory theories
folk psychology, 1:358–359
knowledge acquisition, 2:450
self-knowledge, 2:677
Theory view, of categories, 1:135–136
Theta rhythm, 1:275, 2:698
Thinking, 2:739–742. See also Introspection;
Rationality; Reasoning
argument mapping, 1:50–52
behaviorism and, 1:108–109
concepts and language, 1:176–177
conscious thinking, 1:186–189
current research, 2:741–742
descriptive thought, 1:239–242
gestures and, 1:373
historical perspective, 2:739–740
indexical thought, 1:406–407
inner speech, 1:418–419
mental action, 2:497–498
motivated thinking, 2:528–529
object-dependent thought, 1:240–241, 2:569–571
phenomenology of, 1:188
singular thought, 1:38–39, 240–242
unconscious, 1:188, 2:739
Wason selection task, 2:740–741
Thompson, Erik, 2:610
Thornton, Rosalind, 1:415
Thought. See Thinking
Threaded cognition, 2:545–546
Threat, group, 1:433
Tickles, 1:427
Timbre, 1:386–387
Time
decision making, 1:219–220
estimation of, 2:744–745
flavor, 2:733
hearing, 1:388
indexical thought, 1:407
mind/brain and, 2:515
regret, 2:640
temporal discounting, 1:216
working memory, 2:789
Time perception, 2:742–745
interval timing, 2:743–744
methods and paradigms, 2:744
neurobiological substrates, 2:745
overview, 2:742–743
prospective timing models, 2:744–745
Tinbergen, Niko, 1:189
Titchener, E. B., 1:439
Token identity theory, 2:500
Token-reflexive view, 2:572
Token-token physicalism, 2:518
Tomlin, Michael, 1:62
Tomkins, Silvan, 1:352
Tone deafness, 2:547
Tone languages, 1:371
Tool use, 2:534
Topics model, 2:682
Total time hypothesis, 2:661
Touch, 1:235
Toulmin, Stephen, 1:51
TRACE computational model, 1:157
Traces, memory, 1:144–145
Trading Places (film), 1:148
Tragedy of the commons, 1:160
Trails test, 1:18
Traits, character, 1:147–150
Tranel, Daniel, 1:317
Transcortical motor aphasia, 1:44
Transcortical sensory aphasia, 1:44
Transcranial magnetic stimulation (TMS), 1:123,
289, 316
Transfer of learning, 1:242–243, 251, 2:692–693
Transfer-appropriate processing, 2:661
Transformation theory, 2:582
Transformational models of similarity, 2:690
Translational flow, 2:573–574
Transparency, 1:112, 2:676–677
Treisman, Anne, 2:562, 764
Trial and error learning, 1:393
Triangular subtheory of love, 2:473–474
Trigeminal nerve, 2:732
Tripartite theory of attitude structure, 1:79
Trueswell, John, 2:617
Trunk apraxia, 1:48
Truth conditions, 1:186
Tschermak, Armin, 2:720
Tulving, Endel, 1:191
Tuomela, Raimo, 2:444
Turiel, Elliot, 1:302, 2:526
Turing, Alan, 1:178, 187
Turing machine, 1:187
Turing’s Test, 1:178
Turner, Jason, 1:329
Turner syndrome, 1:347
Tversky, Amos, 1:29, 102, 212, 222, 2:565, 566,
657–658, 689
Twin Earth thought experiment, 1:38–39, 152
Twin studies
<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>heritability</td>
<td>1:390</td>
</tr>
<tr>
<td>substance use disorders</td>
<td>1:10</td>
</tr>
<tr>
<td>Two minds hypothesis</td>
<td>2:741</td>
</tr>
<tr>
<td>Two system models of reasoning</td>
<td>1:209, 229, 2:745–748. See also Dual-process models</td>
</tr>
<tr>
<td>Two-choice reaction time tasks</td>
<td>2:631</td>
</tr>
<tr>
<td>Tye, Michael</td>
<td>1:188, 189</td>
</tr>
<tr>
<td>Tyler, Lorraine</td>
<td>1:156</td>
</tr>
<tr>
<td>Type identity theory</td>
<td>2:500, 638. See also Identity theory; Mind-brain identity theory</td>
</tr>
<tr>
<td>Type-B physicalism</td>
<td>1:286–287</td>
</tr>
<tr>
<td>Uleman, James</td>
<td>2:753</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>1:216</td>
</tr>
<tr>
<td>Unconditioned reinforcers</td>
<td>2:645</td>
</tr>
<tr>
<td>Unconscious, the. See also Consciousness and the unconscious</td>
<td></td>
</tr>
<tr>
<td>Unconscious emotion, psychological perspectives</td>
<td>1:321, 2:749–750</td>
</tr>
<tr>
<td>Unconscious perception</td>
<td>2:751–755</td>
</tr>
<tr>
<td>anesthesia</td>
<td>1:31</td>
</tr>
<tr>
<td>blindsight</td>
<td>1:120–123</td>
</tr>
<tr>
<td>controversies</td>
<td>2:753</td>
</tr>
<tr>
<td>empirical evidence</td>
<td>2:753–754</td>
</tr>
<tr>
<td>historical perspective</td>
<td>2:752–753</td>
</tr>
<tr>
<td>memory</td>
<td>2:752</td>
</tr>
<tr>
<td>processes vs. stimuli</td>
<td>2:751–752</td>
</tr>
<tr>
<td>recent research</td>
<td>2:754</td>
</tr>
<tr>
<td>sensation</td>
<td>2:752</td>
</tr>
<tr>
<td>types of unconsciousness</td>
<td>2:751</td>
</tr>
<tr>
<td>Undertrading</td>
<td>1:322</td>
</tr>
<tr>
<td>Underwood, Benton</td>
<td>2:483</td>
</tr>
<tr>
<td>Unhappiness</td>
<td>1:324. See also Happiness</td>
</tr>
<tr>
<td>Unit weight models of prediction</td>
<td>2:614</td>
</tr>
<tr>
<td>Unitization</td>
<td>1:251</td>
</tr>
<tr>
<td>UNIVAC computer</td>
<td>1:213</td>
</tr>
<tr>
<td>Universal grammar</td>
<td>1:413–414, 417, 2:455</td>
</tr>
<tr>
<td>Universality</td>
<td></td>
</tr>
<tr>
<td>attachment theory</td>
<td>2:648</td>
</tr>
<tr>
<td>child conceptual development</td>
<td>1:170–171</td>
</tr>
<tr>
<td>facial expressions</td>
<td>1:352, 353</td>
</tr>
<tr>
<td>gestures</td>
<td>1:374</td>
</tr>
<tr>
<td>language</td>
<td>1:413–417, 2:455</td>
</tr>
<tr>
<td>University of Wurzburg</td>
<td>2:739</td>
</tr>
<tr>
<td>Up-states</td>
<td>1:275</td>
</tr>
<tr>
<td>U.S. Drug Enforcement Administration's Controlled Substances Act (CSA)</td>
<td>1:8</td>
</tr>
<tr>
<td>U.S. Supreme Court</td>
<td>1:357</td>
</tr>
<tr>
<td>U-shapes developmental curves</td>
<td>2:656</td>
</tr>
<tr>
<td>Utility</td>
<td>1:221–223, 247–249</td>
</tr>
<tr>
<td>Utility of life</td>
<td>1:380</td>
</tr>
<tr>
<td>Utilization behavior</td>
<td>2:595</td>
</tr>
<tr>
<td>Valdesolo, Piercarlo</td>
<td>1:303</td>
</tr>
<tr>
<td>Valence hypothesis</td>
<td>1:288–289, 315</td>
</tr>
<tr>
<td>Validity (research)</td>
<td>2:467, 540</td>
</tr>
<tr>
<td>Valley, George</td>
<td>1:213</td>
</tr>
<tr>
<td>Van de Ven, Niels</td>
<td>1:323</td>
</tr>
<tr>
<td>Van den Broek, Paul</td>
<td>1:250</td>
</tr>
<tr>
<td>Van Dijk, Teun</td>
<td>1:249</td>
</tr>
<tr>
<td>Variable substitution</td>
<td>2:657</td>
</tr>
<tr>
<td>Vector space model</td>
<td>2:682</td>
</tr>
<tr>
<td>Vehicle externalism</td>
<td>1:335</td>
</tr>
<tr>
<td>Velmans, Max</td>
<td>2:612</td>
</tr>
<tr>
<td>Ventral tegmental area</td>
<td>1:215</td>
</tr>
<tr>
<td>Ventral visual stream</td>
<td>2:563</td>
</tr>
<tr>
<td>Ventromedial prefrontal cortex</td>
<td>1:216–217</td>
</tr>
<tr>
<td>Verb tenses</td>
<td>2:460</td>
</tr>
<tr>
<td>Verbal thinking</td>
<td>1:418</td>
</tr>
<tr>
<td>Verhaegen, Paul</td>
<td>1:18</td>
</tr>
<tr>
<td>Veridical experience</td>
<td>1:255–256</td>
</tr>
<tr>
<td>Verificationism</td>
<td>2:517</td>
</tr>
<tr>
<td>Vicary, James</td>
<td>2:724, 753</td>
</tr>
<tr>
<td>Victoria, Queen of England</td>
<td>1:30</td>
</tr>
<tr>
<td>Vigliocco, Gabriella</td>
<td>2:462</td>
</tr>
<tr>
<td>Violence</td>
<td>1:433–434</td>
</tr>
<tr>
<td>Virtue theory</td>
<td>1:147–150, 301</td>
</tr>
<tr>
<td>Vision</td>
<td></td>
</tr>
<tr>
<td>confabulation</td>
<td>1:185</td>
</tr>
<tr>
<td>optic flow</td>
<td>2:573–576</td>
</tr>
<tr>
<td>smell compared to</td>
<td>2:700–701</td>
</tr>
<tr>
<td>visual imagery vs.</td>
<td>2:758–759</td>
</tr>
<tr>
<td>Visual action recognition</td>
<td>1:162</td>
</tr>
<tr>
<td>Visual agnosia</td>
<td>2:766</td>
</tr>
<tr>
<td>Visual imagery</td>
<td>2:757–760, 769. See also Mental representations; Visuospatial reasoning</td>
</tr>
<tr>
<td>characteristics</td>
<td>2:758</td>
</tr>
<tr>
<td>cognitive role</td>
<td>2:759–760</td>
</tr>
<tr>
<td>controversies</td>
<td>2:758</td>
</tr>
<tr>
<td>experimental studies</td>
<td>2:757–758</td>
</tr>
<tr>
<td>subjective qualities</td>
<td>2:757</td>
</tr>
<tr>
<td>vision vs.</td>
<td>2:758–759</td>
</tr>
<tr>
<td>Visual masking</td>
<td>2:760–762. See also Masking</td>
</tr>
<tr>
<td>Visual memory</td>
<td>2:766–767</td>
</tr>
<tr>
<td>Visual neglect</td>
<td>2:578</td>
</tr>
<tr>
<td>Visual working memory</td>
<td>2:765–768</td>
</tr>
<tr>
<td>Visual world paradigm</td>
<td>1:367</td>
</tr>
<tr>
<td>Visual-spatial sketchpad</td>
<td>2:766</td>
</tr>
</tbody>
</table>
Visuospatial reasoning, 2:769–772. See also Perceptual constancy; Spatial cognition, development of; Visual imagery created spaces, 2:771–772
mental transformations, 2:769–770
spatial mental representations, 2:770–771
visual images, 2:769
Visuospatial sketchpad, 2:767, 792
Viviani, Paolo, 1:162
Vocabulary development, 2:778–782
Voice pitch, 1:371. See also Pitch
Voicing, 2:716, 717
Vokey, John, 2:753
Voluntary action, illusion of, 2:772–773. See also Free will; Freedom of action
Von Neumann, John, 1:20, 222, 2:565
Vuilleumier, Patrik, 1:317
Vul, Edward, 2:777

Wada test, 1:288, 315
Wager, Tor, 1:319
Wagner, Allan, 2:645
Wagner, Mark, 2:581, 582
Wainryb, Cecilia, 2:526
Wakeful consciousness, 2:498
Walsh, Vincent, 1:268
Walter, Charles, 2:531
Wants, 1:381
Ward, Geoff, 2:642
Ward, Robert, 1:54
Ward, Ronnie, 1:54
Warren, William, 2:574, 575
Warrington, Elizabeth, 1:400
Wason, Peter, 1:227, 2:739–740
Wason selection task, 1:227, 2:740–741
Waterhouse, Lynn, 2:542
Waterman, Ian, 2:596
Waters, Gloria, 2:794
Watkins, Anthony, 1:386
Watson, John B., 1:105, 108–109, 110, 2:739
Weber, Ernst, 2:743
Weber, Max, 1:158
Weber's law, 1:88, 2:743
Wechsler, David, 1:422
Wechsler Intelligence Scale for Children, 1:355–356, 422
Wegner, Daniel, 2:772
Weinberg, Jonathan, 1:330
Weiskopf, Dan, 1:174
Weiskrantz, Lawrence (Larry), 1:121, 400, 2:754
Welford, Alan, 1:60, 2:623
Wellman, Henry, 1:358
We-mode, 1:158–160
Wernicke, Carl, 1:181, 183

Wernicke-Korsakoff Syndrome (WKS), 1:23
Wernicke's aphasia, 1:44
Wernicke's area, 1:277, 279, 2:456
Whately, Richard, 1:51
Wheatstone, Charles, 2:720
Whitehead, Alfred North, 1:236, 2:513
Whorf, Benjamin, 1:178
Wickens, Christopher, 1:61, 2:544
Wicker, Allen, 1:79
Wide computationalism, 1:335
Wierzbicka, Anna, 1:173, 353
Wigmore, John Henry, 1:51
Wilde, Oscar, 1:102
Wildgruber, Dirk, 1:318
Will. See Free will; Preconscious free will; Voluntary action, illusion of
Williams, Robert, 1:119
Williams syndrome, 2:459
Williams syndrome, 2:775–776
Williamson, Timothy, 2:448
Wilson, Deirdre, 2:511
Wilson, Robert, 1:335
Wilson, Timothy, 1:437, 439
Wisconsin Card Sorting Test, 1:132
Wisdom of crowds effect, 1:257, 374, 2:776–778
Wisniewski, Ed, 1:180
Withdrawal behaviors, 1:288–289
Wittgenstein, Ludwig, 2:597, 676
Wixted, John, 2:481
Wohlschläger, Andreas, 1:162
Woldorff, Marty, 1:59
Wolfe, Jeremy, 2:579, 764
Wolpe, Joseph, 1:105, 109
Woodruff, Guy, 1:168
Woodward, James, 2:635
Word characteristics, 1:115
Word combinations, 1:179–181
Word learning, 2:778–782. See also Language development
adulthood, 2:782
bias, 2:781
brain, 2:781–782
child development, 2:779–781
effects, 2:780–781
overview, 2:778–779
reading, 2:782
Word order languages, 2:618
Word processing, 1:163–165
Word recognition, auditory, 2:783–784
cohort model, 1:156–158
definitions, 1:366–367
frequency effects, 1:366–367
machine speech recognition, 2:477–478

See also entries beginning Dyslexia
Word superiority effect, 2:786
Word-meaning blindness, 1:267
**Working memory, 2:786–791**
  - aging, 1:14–15
  - attention, 1:62–65, 261
  - automaticity, 2:795
  - category learning, 1:130
  - concept of, 1:422
  - emotion, 1:308–310
  - individual and group differences, 2:790–791
  - intelligence, 1:422–423
  - limits, 2:789, 791
  - measurement of, 2:793–794
  - methods of studying, 2:795–796
  - misconceptions, 2:792
  - overview, 2:786–787
  - related concepts, 2:787–788
  - resources, 2:794
  - studies, 2:788–789
  - support systems, 2:767
  - text learning, 2:796
  - theories, 2:789–790
  - visual working memory, 2:765–768

**Working memory, evolution of, 2:791–793**

**Working memory in language processing, 2:793–796**

Wray, Alison, 2:548

Wright, Crispin, 2:677
Wright, Jen, 1:329
Wright, Larry, 1:332–333
Wright, Sewall, 1:389
Wu, Jing, 1:180
Wundt, Wilhelm, 1:14, 299, 439, 2:549, 743
Wurtz, Robert, 2:575
Wyble, Brad, 1:73
Wyer, Robert, Jr., 2:705
Wynn, Thomas, 2:792–793

Yaniv, Ilan, 2:778
Yates, Andrea, 2:628
Young, Andrew, 1:127–128, 231
Young, Andy, 1:350
Young, Garry, 1:128
Young, Jeffrey, 1:107
Young, Liane, 1:304, 329

Zack, Rose, 1:18
Zajonc, Robert, 1:354, 2:754
Zeijlstra, Hedde, 1:415
Ziegler, Jo, 1:268
Zizzo, Daniel, 1:324
Zombies, 1:122
Zwitserlood, Pienie, 1:157