SAGE collegepublishing

Teaching isn't easy. | **Learning** never ends.
We are here for you.

Learn more about SAGE teaching and learning solutions for your course at sagepub.com/collegepublishing.

About SAGE

Founded in 1965, SAGE is a leading independent academic and professional publisher of innovative, high-quality content. Known for our commitment to quality and innovation, SAGE has helped inform and educate a global community of scholars, practitioners, researchers, and students across a broad range of subject areas.

Cover image: iStock.com/Just_Super

ISBN 978-1-5443-8015-5

SAGE college publishing

FRIEDENBERG / SILVERMAN / SPIVEY

COGNITIVE SCIENCE

FOURTH EDITION

SAGE

JAY FRIEDENBERG / GORDON SILVERMAN / MICHAEL JAMES SPIVEY

COGNITIVE SCIENCE

AN INTRODUCTION TO THE STUDY OF MIND

FOURTH EDITION



Cognitive Science

Fourth Edition

Sara Miller McCune founded SAGE Publishing in 1965 to support the dissemination of usable knowledge and educate a global community. SAGE publishes more than 1000 journals and over 600 new books each year, spanning a wide range of subject areas. Our growing selection of library products includes archives, data, case studies and video. SAGE remains majority owned by our founder and after her lifetime will become owned by a charitable trust that secures the company's continued independence.

Los Angeles | London | New Delhi | Singapore | Washington DC | Melbourne

Cognitive Science

An Introduction to the Study of Mind

Fourth Edition

Jay Friedenberg
Manhattan College

Gordon Silverman
Manhattan College

Michael J. Spivey
University of California-Merced

SAGE

Los Angeles | London | New Delhi
Singapore | Washington DC | Melbourne



FOR INFORMATION:

SAGE Publications, Inc.
2455 Teller Road
Thousand Oaks, California 91320
E-mail: order@sagepub.com

SAGE Publications Ltd.
1 Oliver's Yard
55 City Road
London EC1Y 1SP
United Kingdom

SAGE Publications India Pvt. Ltd.
B 1/I 1 Mohan Cooperative Industrial Area
Mathura Road, New Delhi 110 044
India

SAGE Publications Asia-Pacific Pte. Ltd.
18 Cross Street #10-10/11/12
China Square Central
Singapore 048423

Sponsoring Editor: Jessica Miller
Project Associate: Ivey Mellem
Production Editor: Astha Jaiswal
Copy Editor: Gillian Dickens
Typesetter: C&M Digitals (P) Ltd.
Cover Designer: Candice Harman
Marketing Manager: Victoria Velasquez

Copyright © 2022 by SAGE Publications, Inc.

All rights reserved. Except as permitted by U.S. copyright law, no part of this work may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without permission in writing from the publisher.

All third-party trademarks referenced or depicted herein are included solely for the purpose of illustration and are the property of their respective owners. Reference to these trademarks in no way indicates any relationship with, or endorsement by, the trademark owner.

Printed in the United States of America

ISBN 9781544380155

This book is printed on acid-free paper.

21 22 23 24 25 10 9 8 7 6 5 4 3 2 1

BRIEF TABLE OF CONTENTS

Preface xix

About the Authors xxvii

CHAPTER 1 • Introduction: Exploring Mental Space 1

CHAPTER 2 • The Philosophical Approach: Enduring Questions 25

CHAPTER 3 • The Psychological Approach: A Profusion of Theories 57

CHAPTER 4 • The Cognitive Approach I: Vision, Pattern Recognition, and Attention 83

CHAPTER 5 • The Cognitive Approach II: Memory, Imagery, Concepts, and Problem Solving 111

CHAPTER 6 • The Neuroscience Approach: Mind as Brain 147

CHAPTER 7 • The Network Approach: Mind as a Web 189

CHAPTER 8 • The Evolutionary Approach: Change Over Time 229

CHAPTER 9 • The Linguistic Approach: Language and Cognitive Science 263

CHAPTER 10 • The Emotional Approach: Mind as Emotion 297

CHAPTER 11 • The Social Approach: Mind as Society 319

CHAPTER 12 • The Artificial Intelligence Approach I: The Computer as a Cognitive Agent 355

CHAPTER 13 • The Artificial Intelligence Approach II: Embedded Intelligence and Robotics 401

CHAPTER 14 • The Embodied Ecological Approach: A Dynamic Future for Cognitive Science? 437

Glossary 475

References 495

Index 529

DETAILED TABLE OF CONTENTS

Preface xix

About the Authors xxvii

CHAPTER 1 • Introduction: Exploring Mental Space 1

A Brave New World 1
What Is Cognitive Science? 2
Representation 3
Types of Representation 5
Computation 7
The Tri-Level Hypothesis 8
Differing Views of Representation and Computation 10
The Interdisciplinary Perspective 12
The Philosophical Approach 14
INTERDISCIPLINARY CROSSROADS: Science and Philosophy 14
The Psychological Approach 15
The Cognitive Approach 16
The Neuroscience Approach 17
The Network Approach 18
The Evolutionary Approach 18
The Linguistic Approach 19
The Emotion Approach 19
The Social Approach 20
The Artificial Intelligence Approach 20
The Robotics Approach 21
The Embodied Ecological Approach 21
Integrating Approaches 22
Summing Up: A Review of Chapter 1 22

CHAPTER 2 • The Philosophical Approach: Enduring Questions 25

What Is Philosophy? 25
Chapter Overview 25
The Mind–Body Problem: What Is Mind? 26

Monism 28
Evaluating the Monist Perspective 29
Dualism 30
Substance Dualism 31
Property Dualism 32
Evaluating the Dualist Perspective 32
Functionalism: Are Minds Limited to Brains? 34
Evaluating the Functionalist Perspective 36
The Knowledge Acquisition Problem: How Do We Acquire Knowledge? 37
Evaluating the Knowledge Acquisition Debate 39
The Mystery of Consciousness: What Is Consciousness and How Does It Operate? 41
The What-It's-Like Argument 42
Mind as an Emergent Property 44
Evaluating the Emergent View of Mind 46
Consciousness: One or Many? 46
Consciousness and Neuroscience 49
INTERDISCIPLINARY CROSSROADS: Philosophy, Neuroscience, and Binocular Rivalry 51
Consciousness and Artificial Intelligence 53
Overall Evaluation of the Philosophical Approach 55
Summing Up: A Review of Chapter 2 56

CHAPTER 3 • The Psychological Approach: A Profusion of Theories 57

What Is Psychology? 57
Psychology and the Scientific Method 58
Intelligence Tests 61
Mental Atoms, Mental Molecules, and a Periodic Table of the Mind: The Voluntarist Movement 62
Evaluating the Voluntarist Approach 64
Structuralism: What the Mind Is 65
Evaluating the Structuralist Approach 66
Functionalism: What the Mind Does 66
Evaluating the Functionalist Approach 68
The Whole Is Greater Than the Sum of Its Parts: Mental Physics and the Gestalt Movement 69
INTERDISCIPLINARY CROSSROADS: Gestalt Phenomenology, Experimental Psychology, and Perceptual Grouping 71
Evaluating the Gestalt Approach 74
Mini Minds: Mechanism and Psychoanalytic Psychology 75
Evaluating the Psychoanalytic Approach 77

Mind as a Black Box: The Behaviorist Approach 78
Evaluating the Behaviorist Approach 80
Overall Evaluation of the Psychological Approach 81
Summing Up: A Review of Chapter 3 81

CHAPTER 4 • The Cognitive Approach I: Vision, Pattern Recognition, and Attention 83

Some History First: The Rise of Cognitive Psychology 83
The Cognitive Approach: Mind as an Information Processor 84
Modularity of Mind 85
Evaluating the Modular Approach 85
Theories of Vision and Pattern Recognition: How Do We Recognize Objects? 86
Template Matching Theory 87
Evaluating Template Matching Theory 87
Feature Detection Theory 88
Evaluating Feature Detection Theory 90
Recognition by Components Theory 91
Evaluating Recognition by Components Theory 93
INTERDISCIPLINARY CROSSROADS: Computational Vision and Pattern Recognition 94
Evaluating Marr's Computational Approach to Vision 96
Feature Integration Theory 96
Evaluating Feature Integration Theory 100
Theories of Attention: How Do We Pay Attention? 100
Broadbent's Filter Model 101
Evaluating the Filter Model 103
Treisman's Attenuation Model 103
The Deutsch–Norman Memory Selection Model 103
The Multimode Model of Attention 104
Kahneman's Capacity Model of Attention 105
Evaluating the Capacity Model of Attention 107
Evaluating the Model-Building Approach 107
Summing Up: A Review of Chapter 4 108

CHAPTER 5 • The Cognitive Approach II: Memory, Imagery, Concepts, and Problem Solving 111

Types of Memory: How Do We Remember? 111
Sensory Memory 112
Working Memory 114
Scanning Items in Working Memory 117
Long-Term Memory 118
Memory Models 121

The Modal Model 121
Evaluating the Modal Model 122
The Working Memory Model 123
Evaluating the Working Memory Model 125
Visual Imagery: How Do We Imagine? 125
The Kosslyn and Schwartz Theory of Visual Imagery 126
Image Structures 126
Image Processes 127
Evaluating the Kosslyn and Schwartz Theory 129
Concepts: How Do We Represent Our Knowledge of Concepts? **132**
Problem Solving: How Do We Solve Problems? **136**
The General Problem Solver Model 139
Evaluating the General Problem Solver Model 140
INTERDISCIPLINARY CROSSROADS: Artificial Intelligence, Problem Solving, and the SOAR Model **141**
Evaluating the SOAR Model 143
Overall Evaluation of the Cognitive Approach **144**
Summing Up: A Review of Chapter 5 **144**

CHAPTER 6 • The Neuroscience Approach: Mind as Brain **147**

The Neuroscience Perspective **147**
Methodology in Neuroscience **148**
Techniques for the Study of Brain Damage 148
Evaluating Techniques for the Study of Brain Damage 148
Brain Recording Methods 149
Positron Emission Tomography 150
Functional Magnetic Resonance Imaging 150
Magnetoencephalography 151
Knife-Edge Scanning Microscope 152
Brain Stimulation Techniques 152
Electrode Stimulation 152
Transcranial Magnetic Stimulation 152
Optogenetics 153
The Small Picture: Neuron Anatomy and Physiology **153**
The Big Picture: Brain Anatomy **156**
Directions in the Nervous System 157
The Cortex 157
The Split Brain 159
The Neuroscience of Visual Object Recognition **160**
Visual Agnosias 161
Apperceptive Agnosia 162
Associative Agnosia 163
Face Perception 165

INTERDISCIPLINARY CROSSROADS: Perceptual Binding and Neural Synchrony 167
The Neuroscience of Attention 168
Neural Models of Attention 170
A Component Process Model 170
Distributed Network Models 172
Disorders of Attention 173
Hemispatial Neglect 173
Attention-Deficit Hyperactivity Disorder 174
The Neuroscience of Memory 175
Learning and Memory 175
The Hippocampal System 176
Neural Substrates of Working Memory 178
Evaluating the Neuroscience of Working Memory 180
Neural Substrates of Long-Term Memories 180
The Neuroscience of Executive Function and Problem Solving 181
Theories of Executive Function 183
Overall Evaluation of the Neuroscience Approach 186
Summing Up: A Review of Chapter 6 186

CHAPTER 7 • The Network Approach: Mind as a Web 189
The Network Perspective 189
Artificial Neural Networks 190
Characteristics of Artificial Neural Networks 193
Early Conceptions of Neural Networks 194
Backpropagation 195
NETtalk: An Example of a Backpropagation Artificial Neural Network 197
The Elman Net: An Example of a Simple Recurrent Network 199
Evaluating the Connectionist Approach 200
Advantages 200
Problems and Disadvantages 201
Semantic Networks: Meaning in the Web 203
Characteristics of Semantic Networks 203
A Hierarchical Semantic Network 205
Evaluating the Hierarchical Semantic Network Model 207
Evaluating Semantic Networks 208
Network Science 211
Centrality 212
Hierarchical Networks and the Brain 212
Small-World Networks: It's a Small World After All 213
Ordered and Random Connections 215
Egalitarians and Aristocrats 217

Neuroscience and Networks 217
Small-World Networks and Synchrony 219
Percolation 220
Percolation and Psychology 220
The Future of Network Science 222
Overall Evaluation of the Network Approach 222
INTERDISCIPLINARY CROSSROADS: Emotions and Networks 223
Summing Up: A Review of Chapter 7 226

CHAPTER 8 • The Evolutionary Approach: Change Over Time 229

The Evolutionary View 229
A Little Background: Natural Selection and Genetics 230
Comparative Cognition 232
Cognitive Adaptation in Animals 233
INTERDISCIPLINARY CROSSROADS: Evolutionary Processes and Artificial Life 235
Comparative Neuroscience 237
Evaluating the Comparative Approach 240
Evolutionary Psychology 241
Evolved Psychological Mechanisms 242
Evolution and Cognitive Processes 244
Categorization 244
Memory 245
Logical Reasoning 247
Judgment Under Uncertainty 249
Language 252
Behavioral Economics: How We Think About Profit and Loss 253
Sex Differences in Cognition 255
Evaluating Evolutionary Psychology 258
Overall Evaluation of the Evolutionary Approach 261
Summing Up: A Review of Chapter 8 261

CHAPTER 9 • The Linguistic Approach: Language and Cognitive Science 263

The Linguistic Approach: The Importance of Language 263
The Nature of Language 264
Language Processing 265
Phonology 265
Morphology 268
Word Recognition 269
Syntax 270
Semantics 274
Pragmatics 276

Language Acquisition 277
Domain-General and Domain-Specific Mechanisms in Language Acquisition 279
Evaluating Language Acquisition 281
Language Deprivation 282
Evaluating Language Deprivation 284
INTERDISCIPLINARY CROSSROADS: Language, Philosophy, and the Linguistic Relativity Hypothesis 285
Evaluating the Linguistic Relativity Hypothesis 286
Language Use in Nonhuman Animals 286
Evaluating Language Use in Nonhuman Animals 288
Neuroscience and Linguistics: The Wernicke–Geschwind Model 289
Evaluating the Wernicke–Geschwind Model 292
Artificial Intelligence and Linguistics: Natural Language Processing 293
Computer Language Programs and IBM's Watson 293
Evaluation of Natural Language Processing 294
Overall Evaluation of the Linguistic Approach 294
Summing Up: A Review of Chapter 9 295

CHAPTER 10 • The Emotional Approach: Mind as Emotion 297
Emotion and Cognitive Science 297
What Is Emotion? 297
Theories of Emotion 298
Basic Emotions 299
Emotions, Evolution, and Psychological Disorders 300
Disgust 302
Fear 302
Anger 302
Sadness 303
Happiness 303
Emotions and Neuroscience 304
The Chemical and Electrical Basis of Emotional Computation 305
Hot and Cold: Emotion–Cognition Interactions 306
Emotion and Perception/Attention 307
Emotion and Memory 308
Emotion, Mood, and Memory 309
Emotion and Decision Making 310
Emotions and Reasoning by Analogy 311
Emotions and Artificial Intelligence: Affective Computing 312
INTERDISCIPLINARY CROSSROADS: Emotion, Robotics, and the Kismet Project 314
Overall Evaluation of the Emotional Approach 317
Summing Up: A Review of Chapter 10 317

CHAPTER 11 • The Social Approach: Mind as Society 319

Social Cognition 319
Social Cognitive Neuroscience 321
Topics in Social Cognitive Neuroscience 322
- Evolution 322
- Attention 323
- Mirror Neurons 325
- Social Cognition as the Brain's Default State 327

Is Social Cognitive Neuroscience Special? 328
Advantages of the Social Cognitive Neuroscience Approach 329
Theory of Mind 329
- ToM and Neuroscience 330
- Autism 332
- Autism and ToM 333

Other Social Cognitive Disorders 334
Attitudes 334
- Cognitive Dissonance 336
- Attitudes and Cognitive Processes 337
  - *Perception* 337
  - *Attention* 338
  - *Interpretation* 338
  - *Learning* 338
  - *Memory* 338
- Attitudes and Neuroscience 339

Impressions 340
- The Dual-Process Model of Impression Formation 340

Attribution 341
- Attribution Biases 342
- Attribution and Cognitive Processes 342
- Attribution and Neuroscience 343
- **INTERDISCIPLINARY CROSSROADS: Game Theory and the Prisoner's Dilemma** 345

Stereotypes 347
- Stereotypes and Cognitive Processes 347
- Ingroups and Outgroups 348
- Automatic Stereotyping 348
- Stereotyping and Neuroscience 349

Prejudice 350
- The Stereotype Content Model of Prejudice 350

Overall Evaluation of the Social Approach 351
Summing Up: A Review of Chapter 11 352

CHAPTER 12 • The Artificial Intelligence Approach I: The Computer as a Cognitive Entity 355

Cognitive Science and Artificial Intelligence 355
Defining AI 358
Practical AI 360
Introduction 361
AI Implementation ("Hardware") 363
Information and Intelligent Agents 366
Logic, Classical and Fuzzy 367
Reasoning Modalities 367
The Legacies of Turing and Zadeh 368
Turing's State Variable Approach 368
Fuzzy Problem Solving 371
Intelligent Agents That Think, Learn, and Make Decisions 372
Fundamental Concepts of the IA 373
Basic Models 375
Machine Learning and the Data Sciences 377
The Trial-and-Error Learning Method 377
A Simple Algorithmic Method (Regression Line Algorithm) 378
Data Clustering 378
Reconsideration of Data Sources 380
Deep Learning (DL) 382
DL Software 383
Learning Experiences 384
Artificial General Intelligence 386
AGI Problems 388
Reverse Engineering the Brain 389
Methodologies 390
The "Organic Brain" and Wetware 392
Assessment of AI and AGI 395
Summing Up: A Review of Chapter 12 400

CHAPTER 13 • The Artificial Intelligence Approach II: Embedded Intelligence and Robotics 401

Mechanical Beginnings 403
Embodied Cognitive Science 403
The Design of Intelligent Robots as "Biologically Inspired" 403
The Importance of Biology 404
Robotic Embodied Intelligence 405
Defining and Describing a Robot 405

The Intelligent Agent Paradigm 407
Properties of an Autonomous Entity 408
Environments of Intelligent Agents 409
A Simple yet "Sophisticated" Robot 410
Evolutionary Embodiments: The Merger of Human Cognitive Behavior, Biology, and Intelligent Agents 411
Challenges to DL and Its Algorithms 412
Emerging Tools 413
Evolutionary Learning Algorithms and Intelligent Agents 414
Evolutionary Computation 415
The Evolutionary Mutation Process 416
Robotic EA Examples 418
Robotic Embodiments 419
Robotic Realizations 421
Cooperating Intelligent Agents and Swarming 427
Swarming Robotics 427
Particle Robots: An Emerging Technology 428
Embedded Intelligence as an Emotion Machine 430
Machine–Human Interactions 431
Brain Waves 432
The Plasticity of the Human Brain 433
Machines Can Teach Humans 433
Overall Evaluation of Embedded Intelligence 434
Summing Up: A Review of Chapter 13 436

CHAPTER 14 • The Embodied Ecological Approach: A Dynamic Future for Cognitive Science? 437

Embodied and Extended Cognition 437
Perceptual Symbol Systems and Motor Affordances 438
Perceptual Simulations 442
Evaluating Embodied Cognition 445
Dynamical Systems Theory 446
Nonlinearity 446
Predictability 447
State Space and Trajectories 448
Attractors 450
Dynamical Representation 451
INTERDISCIPLINARY CROSSROADS: Multiple Approaches to Levels of Explanation in Cognitive Science 452
Dynamical Versus Classical Cognitive Science 454
The Continuity of Mind 454
Modularity Versus Distribularity 455
Component-Dominant Versus Interaction-Dominant Dynamics 455

Internalism Versus Externalism 456
Amodal Versus Modal Representations 457
Feed-Forward Versus Recurrent Pathways 457
Evaluating the Dynamical Perspective 458
Ecological and Extended Cognition **459**
Ecological Perception 459
Sensorimotor Interaction 462
Extended Cognition 464
Evaluating Ecological and Extended Cognition 466
Integrating Cognitive Science **467**
Integration Across Disciplines 467
Integration Across Levels of Description 468
Integration Across Methodologies 469
Integration Across Cultural Differences 469
The Benefits of Cognitive Science **471**
The Future **472**
Summing Up: A Review of Chapter 14 **472**

Glossary **475**
References **495**
Index **529**

PREFACE

One of the most challenging mysteries remaining to science is the human mind. The brain, which serves as the core engine of the mind, is the most complex object in the universe. It is made up of billions of cells sending signals back and forth to each other over trillions of connections. How can we make sense of all this? Recent years have seen great strides in our understanding, and this has been due in part to developments in technology. In this book, we provide an up-to-date introduction to the study of the mind, examining it from an interdisciplinary perspective. We attempt to understand the mind from the perspective of different fields. Among these are philosophy, psychology, neuroscience, networks, evolution, emotional and social cognition, linguistics, artificial intelligence, robotics, and the new framework of embodied cognition. Beyond this, we make attempts to bridge some of these fields, showing what research at the intersection of these disciplines is like. Each chapter in this text is devoted to a particular disciplinary approach and examines the methodologies, theories, and empirical findings unique to each. Come with us as we explore the next great frontier—our inner world.

WHAT'S NEW IN THIS EDITION

For this fourth edition, new content has been added throughout. In Chapter 1 (Introduction), the treatment of formal logic and production systems has been more richly elaborated with concrete examples. Also, a summary of the new Embodied Ecological Approach in Chapter 14 has been included. In Chapter 2 (The Philosophical Approach), a more in-depth exploration of syllogistic reasoning has been added, along with a more detailed discussion of reductionism and how it contrasts with emergence. Also, the discussion of Searle's Chinese room thought experiment has been expanded. In Chapter 3 (The Psychological Approach), a discussion of intelligence tests has been added. In Chapter 4 (The Cognitive Approach I), the description of Anne Treisman's feature integration theory of visual attention was expanded, along with some discussion of Desimone and Duncan's biased competition account. In Chapter 5 (The Cognitive Approach II), a section on conceptual representation has been added. In Chapter 6 (The Neuroscience Approach), the descriptions of various brain-recording methods have been expanded, and discussions of the somatosensory homunculus and of sparse distributed coding were added. In Chapter 7 (The Network Approach), treatments of Elman's simple recurrent network and McClelland and Rogers's connectionist model of category knowledge have been added, along with an expanded discussion of pattern completion. In Chapter 8 (The Evolutionary Approach), a discussion of foraging skills in animals has been included, and the treatment of gender differences in spatial abilities has

been expanded. Chapter 9 (The Linguistic Approach) has been substantially rearranged. The role of linguistics (and Chomsky in particular) in the formation of cognitive science is emphasized early on. The treatment of phonology, morphology, syntax, semantics, and language acquisition has been expanded significantly. Also, sections on spoken word recognition and cognitive linguistics have been added. In Chapter 10 (The Emotional Approach), Lisa Feldman Barrett's proposal that emotions are not discrete but can partially overlap one another has been added, as well as a discussion of how color perception can influence affect. In Chapter 11 (The Social Approach), the discussions of the mirror neuron system, autism, the prisoner's dilemma, and stereotype formation have been slightly expanded. Chapter 12 (The Artificial Intelligence Approach I) has been substantially rearranged in its treatment of Alan Turing and Lotfi Zadeh, with a new focus on intelligent agents, Bayesian probability, deep learning, and brain–computer interfaces. In Chapter 13 (The Artificial Intelligence Approach II), the discussion of reactive and deliberative architectures has been expanded and sections have been added on robotic embodied intelligence, evolutionary algorithms, and swarming robotics. Chapter 14 (The Embodied Ecological Approach) was converted from a "future-looking conclusion" chapter into a "state-of-the-art dynamical, embodied, ecological" chapter. The discussions of dynamical systems theory and ecological perception have been significantly expanded, and a section on embodied cognition has been added.

A Matrix for Exploring Cognitive Science Across Disciplines

Chapter No.	Name/Title	Chapter Summary	Primary Topic/ Issues	Secondary Topic/ Issues	Methodologies	Major Figures	Evaluation
1	Introduction	An introduction to cognitive science and summary overview of different perspectives	• Interdisciplinary study • Representation and computation • Interdisciplinary perspective • Categories of mental representation	• Concepts • Propositions • Production rules • Declarative and procedural knowledge • Analogies	• No methodologies discussed	• Thagard • Harnish • Pylyshyn • Marr	• Cognitive science is unique in that it binds together different perspectives and methodologies in the study of mind
2	The Philosophical Approach	The search for wisdom and knowledge; frames broad questions about mind	• The mind–body problem • Functionalism • Knowledge acquisition • Consciousness	• Monism • Dualism • Nature–nurture debate • Reductionism • Emergence	• Deductive and inductive reasoning	• Aristotle • Plato • Berkeley • Democritus • Descartes • Ryle • Clark • Hume • Locke • Chalmers • Nagel • Jackson • Searle • Churchland • Dennett	• Provides a broad perspective; asks fundamental questions; not an empirical approach
3	The Psychological Approach	The scientific study of mind and behavior	• The scientific method • Voluntarism • Structuralism • Functionalism • Gestalt psychology • Psychoanalytic psychology • Behaviorism	• Theory and hypothesis • Independent and dependent variables • Experimental and control groups • Stream of consciousness • Levels of consciousness	• Scientific method • Introspection • Phenomenology	• Wundt • Titchener • James • Wertheimer • Koffka • Kohler • Freud • Watson • Pavlov • Skinner	• Multiple theoretical positions; first systematic and scientific study of mental phenomena; problems with introspection and phenomenology

(Continued)

(Continued)

Chapter No.	Name/Title	Chapter Summary	Primary Topic/ Issues	Secondary Topic/ Issues	Methodologies	Major Figures	Evaluation
4	The Cognitive Approach I	The information-processing view of mind; use of a computer as a metaphor for mind; use of process models and assumption of modularity	• Information-processing perspective • Modularity • Pattern recognition • Attention	• Template matching • Feature detection • Computational vision • Feature integration theory • Models of attention	• Experimentation • Modeling	• Neisser • Fodor • Selfridge • Norman • Marr • Treisman • Broadbent • Deutch • Posner • Snyder • Kahneman • Biederman	• Fruitful synergistic use of experimentation and model building
5	The Cognitive Approach II	The information-processing view of mind; use of a computer as a metaphor for mind; use of process models and assumption of modularity (Same as Cognitive Approach I chapter)	• Memory • Models of memory • Visual imagery • Problem solving	• Memory types: sensory, working, and long term • The modal, and working memory models • The Kosslyn–Schwartz theory of visual imagery • Heuristics • Means-ends analysis • The GPS and SOAR models	• Experimentation • Modeling • Same as Cognitive Approach I chapter	• Sperling • Baddeley • Atkinson • Shiffrin • Anderson • Kosslyn • Block • Newell • Simon • Sternberg	• Common set of assumptions underlying information processing and modularity; concepts of representation and computation need to be reconciled with connectionism
6	The Neuroscience Approach	The study of nervous system anatomy and physiology that underlies and gives rise to cognitive function	• Neuroscience methodology • Neuron anatomy and physiology • Brain anatomy • Neuroscience of visual object recognition, attention, memory, executive function, and problem solving	• The split brain • Dorsal and ventral pathways • Agnosias • Plasticity • Hippocampal function • Action schemas and scripts • Metacognition • Binding and neural synchrony	• Case studies • Lesion studies • Cell-recording techniques • EEG, ERP, CAT, PET, and fMRI • MEG andTMS	• Sperry • Sacks • Humphreys • Posner • Lashley • Hebb • Shallice • Engel • Singer	• The marriage of cognitive and neuroscience perspectives in cognitive neuroscience is a good integrative approach; specification of biological structures and processes of cognitive abilities

Chapter No.	Name/Title	Chapter Summary	Primary Topic/ Issues	Secondary Topic/ Issues	Methodologies	Major Figures	Evaluation
7	The Network Approach	View of mind as an interconnected set of nodes or web; processing consists of the spread of activation through the web	• Serial and parallel processing • Artificial neural networks • Semantic networks • Network science	• Perceptrons • Back propagation • Stability and plasticity • Catastrophic interference • Spreading activation • Retrieval cues • Priming • Propositional networks • Small-world networks	• Software simulations of artificial neural networks • Comparison of results with theory and empirical data	• McCulloch • Pitts • Hopfield • Kohonen • Grossberg • Collins • Quillian • Rumelhart • McClelland • Watts • Strogatz • Buchanan	• Significant advantages to using networks for understanding learning and knowledge representation; challenges in building networks that rival the brain
8	The Evolutionary Approach	Mind as the adapted product of selection forces	• Natural selection • Evolved psychological mechanisms • Comparative cognition • Evolution and cognitive processes • Behavioral economics • Artificial life	• General-purpose versus domain-specific view of mind • Wason selection task • Heuristics and fallacies • Exaptation, molecular drive, and spandrels	• Experimentation • Cross-species comparison	• Darwin • Buss • Cosmides • Tooby • Edelman	• Powerful theoretical framework, but not all mental processes may be adaptive; good integration with neuroscience; domain-specific processing view clashes with general-purpose processor view

(Continued)

(Continued)

Chapter No.	Name/Title	Chapter Summary	Primary Topic/ Issues	Secondary Topic/ Issues	Methodologies	Major Figures	Evaluation
9	The Linguistic Approach	The multidisciplinary study of language	• The nature of language • Primate language use • Language acquisition • Language deprivation • Linguistic relativity hypothesis • Grammar • The Wernicke–Geschwind model • Natural language processing	• Phonology and morphology • Syntax and semantics • Animal language studies • Critical period • Second-language acquisition • Phrase structure, transformational and universal grammar • Aphasias • Speech recognition • Pragmatic analysis	• Case studies • Developmental studies • Experimentation	• Gardner • Premack • Savage-Rumbaugh • Sapir • Whorf • Chomsky • Broca • Wernicke • Tanenhaus • Clark • Lakoff • Marslen-Wilson • Saffran	• Multiple perspectives and techniques brought to bear on the complex cognitive topic of language; advances in computer-based language comprehension
10	The Emotional Approach	Emotions and moods influence all major cognitive processes and should be incorporated into cognitive theories and models of mind	• Emergence synthesis approach • Flashbulb and autobiographical memory • Mood-congruent and mood-dependent memory • Role of synapses in emotional processing • Affective computing	• Evolutionary accounts of psychological disorders • Analytical rumination hypothesis • The feel-good, do-good phenomenon • CogAff architecture • Social robots	• Experimentation • The Iowa Gambling Task	• Damasio • Fox • Minsky • Solomon	• Need for an understanding of information flow and neural circuits that underlie emotions and how these impact thought; more work needed on how cognitions affect emotions

Chapter No.	Name/Title	Chapter Summary	Primary Topic/ Issues	Secondary Topic/ Issues	Methodologies	Major Figures	Evaluation
11	The Social Approach	Thinking about people is different from thinking about objects; social cognitive neuroscience is a good example of the interdisciplinary approach	• Mirror neurons • Theory of mind • Attitudes • Cognitive dissonance • Impressions • Attributions • Stereotyping • Prejudice	• Anterior and posterior attention systems • Autism • Prisoner's dilemma	• Brain imaging • The ultimatum game • Social dilemmas	• Fiske • Ochsner and Lieberman • Rizzolatti • Siegal and Varley • Frith • Schacter	• Automatic versus effortful processing is a common theme; the historical focus on the individual is too narrow, cognitive science must embrace the social environment
12	The Artificial Intelligence Approach I	Defining the concept of artificial intelligence; machine representation of cognitive function	• Historical perspective • Influence of Turing • Predictive architectures • Artificial general intelligence • Agent-based architectures: Multiagent systems	• Universal computation • Chatbots • Evolutionary computing	• Cognitive models • Turing Test	• Turing • Minsky • Kurzweil • Craik • Hawkins • Russell • Norvig • McCarthy • Goertzel	• New computational technologies may lead to the densities required to achieve the requirements needed to implement an intelligence that is beyond the human level; a fundamental dilemma persists: "Brains must have programs yet at the same time must not be programmed."

(Continued)

(Continued)

Chapter No.	Name/Title	Chapter Summary	Primary Topic/ Issues	Secondary Topic/ Issues	Methodologies	Major Figures	Evaluation
13	The Artificial Intelligence Approach II	The intelligent agent (IA) paradigm; applying the principles to the design of IAs; cognitive models of IAs; robotic embodiments; biological and social influences on robots	• Importance of biology • Embodiment and situational aspects of IA (structure) • Expert systems • Building a brain • Robotic architectures	• Hierarchical, reactive, hybrid robotic architectures • Emotion in IAs	• Cognitive modeling • Simulation	• Turing • Minsky • Brooks • Russell • Norvig • Breazeal • Arkin	• In some activities, machines already outperform humans; a great many problems need to be addressed in the future: perception, finely honed reasoning, and manipulative capabilities of adult humans; the more we try to replicate human intelligence, the more we may learn to appreciate and understand humans
14	The Embodied Ecological Approach	An evaluation of the embodiment approach to cognitive science and the ecological perception framework	• Cognitive science needs to do a better job explaining the role of the body and of physical environments in cognition, as well as individual and cultural differences • The dynamical systems approach	• Predictability • Randomness • Constructivism	• Nonlinear modeling • Use of state space, trajectories, and attractors to describe cognitive phenomena	• Gibson • Dreyfus • Brooks • Barsalou • Pulvermüller • Borghi • Turvey • Spivey	• Benefits of cognitive science are many and widespread throughout engineering, medicine, education, and other fields; lack of a single unified theory

ABOUT THE AUTHORS

Jay Friedenberg is Professor of the Psychology Department at Manhattan College, where he directs the Cognitive Science Program. He is interested in both vision and the philosophy of mind. He teaches courses in physiological psychology, cognition and learning, sensation and perception, and artificial intelligence and robotics. He has published several articles on visual estimation of center of mass. His current research projects focus on the aesthetics of geometrical shapes. He has published books on artificial intelligence, dynamical systems theory, and psychology. He is a member of the International Association of Empirical Aesthetics, the Eastern Psychological Association, the Vision Science Society, the Psychonomic Society, and Phi Beta Kappa. He obtained his PhD in cognitive psychology in 1995 at the University of Virginia.

Gordon Silverman is Professor Emeritus of Electrical and Computer Engineering at Manhattan College. His professional career spans more than 55 years of corporate, teaching, consulting, and research experience, during which he has developed a range of scientific instruments, particularly for use in physiological psychology research environments. He is the holder of eight patents, some related to behavior modification. The author of more than 20 journal articles and books, he has also served on the faculties of The Rockefeller University and Fairleigh Dickinson University. His current research interests include telemedicine, rehabilitation medicine, artificial intelligence, and biomedical instrumentation and modeling. He holds engineering degrees from Columbia University and received a PhD in system science from New York University Polytechnic School of Engineering in 1972.

Michael J. Spivey is Professor of Cognitive Science at the University of California, Merced. He earned his BA in Psychology at the University of California, Santa Cruz, and his PhD in Brain and Cognitive Sciences at the University of Rochester. After 12 years as a psychology professor at Cornell University, Spivey moved to UC Merced to help build their Department of Cognitive and Information Sciences. He has published over 100 journal articles and book chapters on the embodiment of cognition and interactions between language, vision, memory, syntax, semantics, and motor movement. His research uses eye tracking, computer-mouse tracking, and dynamical systems theory to explore how brain, body, and environment work together to make a mind what it is. In 2010, Spivey received the William Procter Prize for Scientific Achievement from the Sigma Xi Scientific Research Honor Society.

CHAPTER ONE

INTRODUCTION

Exploring Mental Space

Learning Objectives

After reading this chapter, you will be able to:

1. List at least five disciplines that participate in the field of cognitive science.
2. Describe what a mental representation is.
3. Describe what mental computation is.
4. Define what interdisciplinary means.

A BRAVE NEW WORLD

We are in the midst of a scientific revolution. For centuries, science has made great strides in our understanding of the external observable world. Physics revealed the motion of the planets, chemistry discovered the fundamental elements of matter, and biology has told us how to understand and treat disease. But during much of this time, there were still many unanswered questions about something perhaps even more important to us—the human mind.

What makes mind so difficult to study is that, unlike the phenomena described above, it is not something we can easily observe, measure, or manipulate. In addition, the mind is the most complex entity in the known universe. To give you a sense of this complexity, consider the following. The human brain is estimated to contain 10 billion to 100 billion individual nerve cells or neurons. Each of these neurons can have as many as 10,000 connections to other neurons. This vast web of neural tissue is the core engine of the mind and helps generate a wide range of amazing and difficult-to-understand mental phenomena, such as perception, memory, language, emotion, and social interaction.

The past several decades have seen the introduction of new technologies and methodologies for studying this intriguing organ, and its relationship to the body, and to the environment. We have learned more about the mind in the past half-century than in all the time that came before that. This period of rapid discovery has coincided with an increase in the number of different disciplines—many of them entirely new—that study mind. Since then, a coordinated effort among the practitioners of these disciplines has come to pass. This diversely interdisciplinary approach has since become known as cognitive science. Unlike the sciences that came before, which were focused solely on the world of physical events in physical space, this new endeavor now turns its full attention to discovering the fascinating mental events that take place in mental space.

WHAT IS COGNITIVE SCIENCE?

Cognitive science can be roughly summed up as the scientific interdisciplinary study of the mind. Its primary methodology is the scientific method—although, as we will see, many other methodologies also contribute. A hallmark of cognitive science is its interdisciplinary approach. It results from the efforts of researchers working in a wide array of fields. These include philosophy, psychology, linguistics, artificial intelligence (AI), robotics, and neuroscience, among others. Each field brings with it a unique set of tools and perspectives. One major goal of this book is to show that when it comes to studying something as complex as the mind, no single perspective is adequate. Instead, intercommunication and cooperation among the practitioners of these disciplines will tell us much more.

The term *cognitive science* refers not so much to the sum of all these disciplines but to their intersection or converging work on specific problems. In this sense, cognitive science is not a unified field of study like each of the disciplines themselves but, rather, a collaborative effort among researchers working in the various fields. The glue that holds cognitive science together is the topic of mind and, for the most part, the use of scientific methods. In the concluding chapter, we talk more about the issue of how unified cognitive science really is.

To understand what cognitive science is all about, we need to know what its theoretical perspective on the mind is. This perspective began with the idea of **computation**, which may alternatively be called information processing. Cognitive scientists started out viewing the mind as an information processor. Information processors must both represent information and transform information. That is, a mind, according to this perspective, must incorporate some form of mental representation and processes that act on and manipulate that information. We will discuss these two ideas in greater detail later in this chapter.

Cognitive science is often described as having been heavily influenced by the development of the digital computer. Computers are, of course, information processors. Think for a minute about a personal computer. It performs a variety of information-processing tasks. Information gets into the computer via input devices, such as a keyboard or modem. That information can then be stored on the computer—for example, on its hard drive—in the form of binary representations (coded as 0s and 1s). The information can then be processed and manipulated using software, such as a text editor. The results of this processing may next serve as output, either to a monitor or to a printer. In like fashion, we may think of people performing similar tasks. Information is "input" into our minds through perception—what we see or hear. It is stored in our memories and processed in the form of thought. Our thoughts can then serve as the basis of "outputs," such as language or physical behavior.

Of course, this analogy between the human mind and computers is highly abstract and imperfect. The actual physical way in which data are stored on a computer bears little resemblance to human memory formation. But both systems are characterized by some form of computation (e.g., binary computation in the case of computers and analog computation in the case of brains). In fact, it is not going too far to say that most

cognitive scientists view the mind as a machine or mechanism whose workings they are trying to understand.

REPRESENTATION

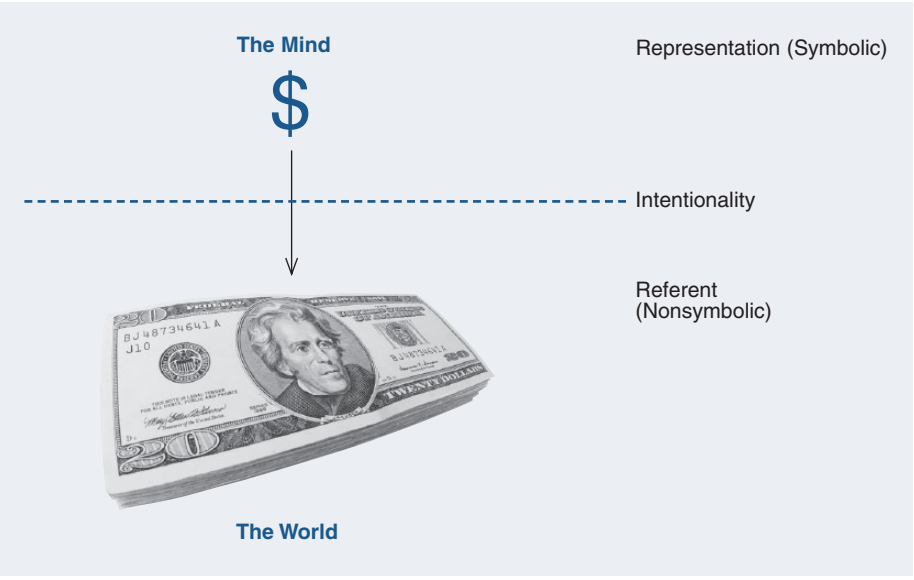
As mentioned before, representation has been seen as fundamental to cognitive science. But what is a representation? Briefly stated, a **representation** is something that stands for something else. Before listing the characteristics of a representation, it is helpful to describe briefly four categories of representation. (1) A concept stands for a single entity or group of entities. Single words are good examples of concepts. The word *apple* refers to the concept that represents that particular type of fruit. (2) Propositions are statements about the world and can be illustrated with sentences. The sentence "Mary has black hair" is a proposition that is itself made up of a few concepts. (3) Rules are yet another form of representation that can specify the relationships between propositions. For example, the rule "If it is raining, I will bring my umbrella" makes the second proposition contingent on the first. (4) An analogy helps us make comparisons between two similar situations. We will discuss all four of these representations in greater detail in the "Interdisciplinary Crossroads" section at the end of this chapter.

There are four crucial aspects of any representation (Hartshorne, Weiss, & Burks, 1931–1958). First, a "representation bearer" such as a human or a computer must realize a representation. Second, a representation must have content—meaning it stands for one or more objects. The thing or things in the external world that a representation stands for are called **referents**. Third, a representation must also be "grounded." That is, there must be some way in which the representation and its referent come to be related. Fourth, a representation must be interpretable by some interpreter, either the representation bearer or somebody else. These and other characteristics of representations are discussed next.

The fact that a representation stands for something else means it is **symbolic**. We are all familiar with symbols. We know, for instance, that the dollar symbol ($) is used to stand for money. The symbol itself is not the money but, instead, is a surrogate that refers to its referent, which is actual money. In the case of mental representation, we say that there is some symbolic entity "in the mind" that stands for real money. Figure 1.1 shows a visual representation of money. Mental representations can stand for many different types of things and are by no means limited to simple conceptual ideas such as "money." Research suggests that there are more complex mental representations that can stand for rules—for example, knowing how to drive a car and analogies, which may enable us to solve certain problems or notice similarities (Thagard, 2000).

Human mental representations, especially linguistic ones, are said to be **semantic**, which is to say that they have meaning. Exactly what constitutes meaning and how a representation can come to be meaningful are topics of debate. According to one view, a representation's meaning is derived from the relationship between the representation and what it is about. The term that describes this relation is **intentionality**. Intentionality means "directed on an object." Mental states and events are intentional. They refer to some actual thing or things in the world. If you think about your brother, then the thought of your brother is directed toward him—not toward your sister, a cloud, or some other object.

Figure 1.1 Different aspects of the symbolic representation of money.



Source: PhotoObjects.net/Thinkstock.

An important characteristic of intentionality has to do with the relationship between inputs and outputs to the world. An intentional representation must be triggered by its referent or things related to it. Consequently, activation of a representation (i.e., thinking about it) should cause behaviors or actions that are somehow related to the referent. For example, if your friend Sally told you about a cruise she took around the Caribbean last December, an image of a cruise ship would probably pop into mind. This might then cause you to ask her if the food onboard was good. Sally's mention of the cruise was the stimulus input that activated the internal representation of the ship in your mind. Once it was activated, it caused the behavior of asking about the food. This relation between inputs and outputs is known as an **appropriate causal relation**.

Symbols can be assembled into what are called physical symbol systems, or more simply as formal logical systems. In a **formal logical system**, symbols are combined into expressions. These expressions can then be manipulated using processes. The result of a process can be a new expression. For example, in formal logic, symbols can be words like *animals* or *mammals*, and expressions could be statements like "animals that nurse their young are mammals." The processes would be the rules of deduction that allow us to derive true *concluding* expressions from *known* expressions. In this instance, we could start off with the two known expressions "animals that nurse their young are mammals," and "whales nurse their young," and generate the new concluding expression: "whales are mammals." (There are, in fact, a few nonmammals that nurse their young, but that is

for an advanced biology textbook.) More on this below, where we discuss propositions and syllogisms.

According to the **physical symbol system hypothesis (PSSH)**, a formal logical system can allow for intelligence (Newell & Simon, 1976). Since we as humans appear to have representational and computational capacity, being able to use things that stand for things, we seem to be intelligent. Beyond this, we could also infer that machines are intelligent, since they too have this capacity, although of course, this is debated.

Several critiques have been leveled against the PSSH (Nilsson, 2007). For example, it is argued that the symbols computers use have no meaning or semantic quality. To be meaningful, symbols have to be connected to the environment in some way. People and perhaps other animals seem to have meaning because we have bodies and can perceive things and act on them. This "grounds" the symbols and imbues them with semantic quality. Computing machines that are not embodied with sensors (e.g., cameras, microphones) and effectors (e.g., limbs) cannot acquire meaning. This issue is known as **the symbol grounding problem** and is in effect a reexpression of the concept of intentionality.

A counterargument to this is that computer systems do have the capability to **designate**. An expression can designate an object if it can affect the object itself or behave in ways that depend on the object. One could argue that a robot capable of perceiving an object like a coffee mug and able to pick it up could develop semantics toward it in the same way that a person might. Ergo, the robot could be intelligent. Also, there are examples of AI programs like expert systems that have no sensor or effector capability yet are able to produce intelligent and useful results. Some expert systems, like MYCIN, are able to more accurately diagnose certain medical disorders than are members of the Stanford University medical school (Cawsey, 1998). They can do this despite not being able to see or act in the world.

Types of Representation

The history of research in cognition suggests that there are numerous forms of mental representation. Paul Thagard (2000), in *Mind: Introduction to Cognitive Science*, proposes four: concepts, propositions, rules, and analogies. Although some of these have already been alluded to and are described elsewhere in the book, they are so central to many ideas in cognitive science that it is, therefore, useful to sketch out some of their major characteristics again here.

A **concept** is perhaps the most basic form of mental representation. A concept is an idea that represents things we have grouped together. The concept "chair" does not refer to a specific chair, such as the one you are sitting in now, but it is more general than that. It refers to all possible chairs no matter what their colors, sizes, and shapes. Concepts need not refer to concrete items. They can stand for abstract ideas—for example, "justice" or "love." Concepts can be related to one another in complex ways. They can be related in a hierarchical fashion, where a concept at one level of organization stands for all members of the class just below it. "Golden retrievers" belongs to the category of "dogs," which in turn belongs to the category of "animals." We discuss a hierarchical model of concept representation in the network approach chapter. The question of whether concepts are innate or learned is discussed in the philosophical approach chapter.

A **proposition** is a statement or assertion typically posed in the form of a simple sentence. An essential feature of a proposition is that it can be proved true or false. For instance, the statement "The moon is made out of cheese" is grammatically correct and may represent a belief that some people hold, but it is a false statement. We can apply the rules of formal logic to propositions to determine the validity of those propositions. One logical inference is called a syllogism. A syllogism consists of a series of propositions. The first two (or more) are premises, and the last is a conclusion. Take the following syllogism:

Premise 1: All men like football.

Premise 2: Charlie is a man.

Conclusion: Charlie likes football.

Obviously, the conclusion can be wrong if either of the two premises is not fully and completely true. If 99% of men like football, then it might not be true that Charlie likes football, even if he is a man. And if Charlie is not a man, then Charlie may or may not like football, even assuming all men like it. Logical conclusions are only as reliable as the premises on which they are based. In the artificial intelligence approach chapter, we will discuss how probabilities (like 99%) can be used in probabilistic reasoning (and even fuzzy logic) to work with premises that are not fully and completely true.

You may have noticed that propositions are representations that incorporate concepts. The proposition "All men like football" incorporates the concepts "men" and "football." Propositions are more sophisticated representations than concepts because they express relationships—sometimes very complex ones—between concepts. The rules of logic are best thought of as computational processes that can be applied to propositions to determine their validity. However, logical relations between propositions may themselves be considered a separate type of representation. The evolutionary approach chapter provides an interesting account of why logical reasoning, which is difficult for many people, can be made easier under certain circumstances.

Formal logic is at the core of a type of computing system that produces effects in the real world: production systems. Inside a production system, a **production rule** is a conditional statement of the following form: "If *x*, then *y*," where *x* and *y* are propositions. In formal logic, the "if" part of the rule is called the antecedent, and the "then" part is called the consequent. In a production system, the "if" part of the rule is called the condition, and the "then" part is called the action. If the proposition that is contained in the condition (*x*) is verified as true, then the action that is specified by the second proposition (*y*) should be carried out, according to the rule. The following rules help us drive our cars:

If the light is red, then step on the brakes.

If the light is green, then step on the accelerator.

In fact, it is production rules like these that are used in the computational algorithms that run self-driving cars. Notice that, in the first rule, the two propositions are "the light is

red" and "step on the brakes." We can also form more complex rules by linking propositions with "and" and "or" statements:

If the light is red or the light is yellow, then step on the brakes.

If the light is green and nobody is in the crosswalk, then step on the accelerator.

The *or* that links the two propositions in the first part of the rule specifies that when *either* proposition is true, the action should be carried out. By contrast, when an *and* links these two propositions, then the rule specifies that *both* must be true before the action can occur.

Rules bring up the question of what knowledge really is. We usually think of knowledge as factual. Indeed, a proposition such as "Candy is sweet," if validated, does provide factual information. The proposition is then an example of declarative knowledge. **Declarative knowledge** is used to represent facts. It tells us what is and is demonstrated by verbal communication. **Procedural knowledge**, by comparison, refers to skills. It tells us how to do something and is demonstrated by action. If we say that World War II was fought during the period 1939 to 1945, we have demonstrated a fact learned in history class. If we ski down a snowy mountain slope in the winter, we have demonstrated that we possess a specific skill. It is, therefore, very important that information-processing systems have some way of representing actions if they are to help an organism or machine perform those actions. Rules are just one way of representing procedural knowledge. In the cognitive approach chapters, we discuss two cognitive, rule-based systems: the atomic components of thought and SOAR (state, operator, and result) models. In the ecological embodied approach chapter, we discuss more sensorimotor ways of representing procedural knowledge (and even declarative knowledge).

Another specific type of mental representation is the **analogy**—although, as is pointed out below, the analogy can also be classified as a form of reasoning. Thinking analogically involves applying one's familiarity with an old situation to a new situation. Suppose you had never ridden on a train before but had taken buses numerous times. You could use your understanding of bus riding to help you figure out how to take a ride on a train. Applying knowledge that you already possess and that is relevant to both scenarios would enable you to accomplish this. Based on prior experience, you would already know that you have to first determine the schedule, perhaps decide between express and local service, purchase a ticket, wait in line, board, stow your luggage, find a seat, and so on.

Analogies are a useful form of representation because they allow us to generalize our learning. Not every situation in life is entirely new. We can apply what we already have learned to similar situations without having to figure out everything all over again. Several models of analogical reasoning have been proposed (Forbus, Gentner, & Law, 1995; Holyoak & Thagard, 1995).

COMPUTATION

As mentioned earlier, representations are only the first key component of the traditional cognitive science view of mental processes. Representations by themselves are of little use

unless something can be done with them. Having the concept of money doesn't do much for us unless we know how to calculate a tip or can give back the correct amount of change to someone. In this cognitive science view, the mind performs computations on representations. It is, therefore, important to understand how these mental mechanisms operate.

What sorts of mental operations does the mind perform? If we wanted to get details about it, the list would be endless. Take the example of mathematical ability. If there were a separate mental operation for each step in a mathematical process, we could say the mind adds, subtracts, divides, and so on. Likewise, with language, we could say that there are separate mental operations for making a noun plural, putting a verb into past tense, and so on. It is better, then, to think of mental operations as falling into broad categories. These categories can be defined by the type of operation that is performed or by the type of information acted on. An incomplete list of these operations would include sensation, perception, attention, memory, language, mathematical reasoning, logical reasoning, decision making, and problem solving. Many of these categories may incorporate virtually identical or similar subprocesses—for example, scanning, matching, sorting, and retrieving. Figure 1.2 shows the kinds of mental processes that may be involved in solving a simple addition problem.

The Tri-Level Hypothesis

Any given information process can be described at several different levels. According to the tri-level hypothesis, biological or artificial information-processing events can be evaluated on at least three different levels (Marr, 1982). The highest or most abstract level of analysis is the **computational level**. At this level, one is concerned with two tasks. The first is a clear specification of what the problem is. Taking the problem as it may originally have been posed, in a vague manner perhaps, and breaking it down into its main constituents or parts can bring about this clarity. It means describing the problem in a precise way such that the problem can be investigated using formal methods. It is like asking, What exactly is this problem? What does this problem entail? The second task one encounters at the computational level concerns the purpose or reason for the process. The second task consists of asking, Why is this process here in the first place? Inherent in this analysis is adaptiveness—the idea that biological mental processes are learned or have evolved to enable the organism to solve a problem it faces. This is the primary explanatory perspective used in the evolutionary approach. We describe a number of cognitive processes and the putative reasons for their evolution in the evolution chapter.

Stepping down one level of abstraction, we can next inquire about the way in which an information process is carried out. To do this, we need an **algorithm**, a formal **procedure** or system that acts on informational representations. It is important to note that algorithms can be carried out regardless of a representation's meaning; algorithms act on the form, not the meaning, of the symbols they transform. One way to think of algorithms is that they are "actions" used to manipulate and change representations. Algorithms are formal, meaning they are well defined. We know exactly what occurs at each step of an algorithm and how a particular step changes the information being acted on. A mathematical formula is a good example of an algorithm. A formula specifies how

Figure 1.2 Some of the computational steps involved in solving an addition problem.



Computational Steps	
1. 6 + 7 = 13	Add right column
2. 3	Store three
3. 1	Carry one
4. 3 + 4 = 7	Add left column
5. 7 + 1 = 8	Add one
6. 8	Store eight
7. 83	Record result

the data are to be transformed, what the steps are, and what the order of steps is. This type of description is put together at the **algorithmic level**, sometimes also called the programming level. It is equivalent to asking, What information-processing steps are being used to solve the problem? To draw an analogy with computers, the algorithmic level is like software because software contains instructions for the processing of data.

The most specific and concrete type of description is formulated at the **implementational level**. Here we ask, What is the information processor made of? What types of physical or material changes underlie changes in the processing of the information? This level is sometimes referred to as the hardware level, since in computer parlance, the hardware is the physical "stuff" the computer is made of. This would include its various parts—a hard drive, a screen, keyboard, and so on. At a smaller scale, computer hardware consists of multiple circuit boards and even the flow of electrons through the circuits. The hardware in biological cognition is the brain and body and, on a smaller scale, the neurons and activities of those neurons.

At this point, one might wonder, Why do we even need an algorithmic or formal level of analysis? Why not just map the physical processes at the implementational level onto a computational description of the problem or, alternatively, onto the behaviors or actions of the organism or device? This seems simpler, and we need not resort to the idea of information and representation. The reason is that the algorithmic level tells us how a particular system performs a computation. Not all computational systems solve a problem in the same way. Both computers and humans can perform addition but do so in drastically different fashions. This is true at the implementational level, obviously, but understanding the difference formally tells us much about alternative problem-solving approaches. It also gives us insights into how these systems might compute solutions to other novel problems that we might not understand.

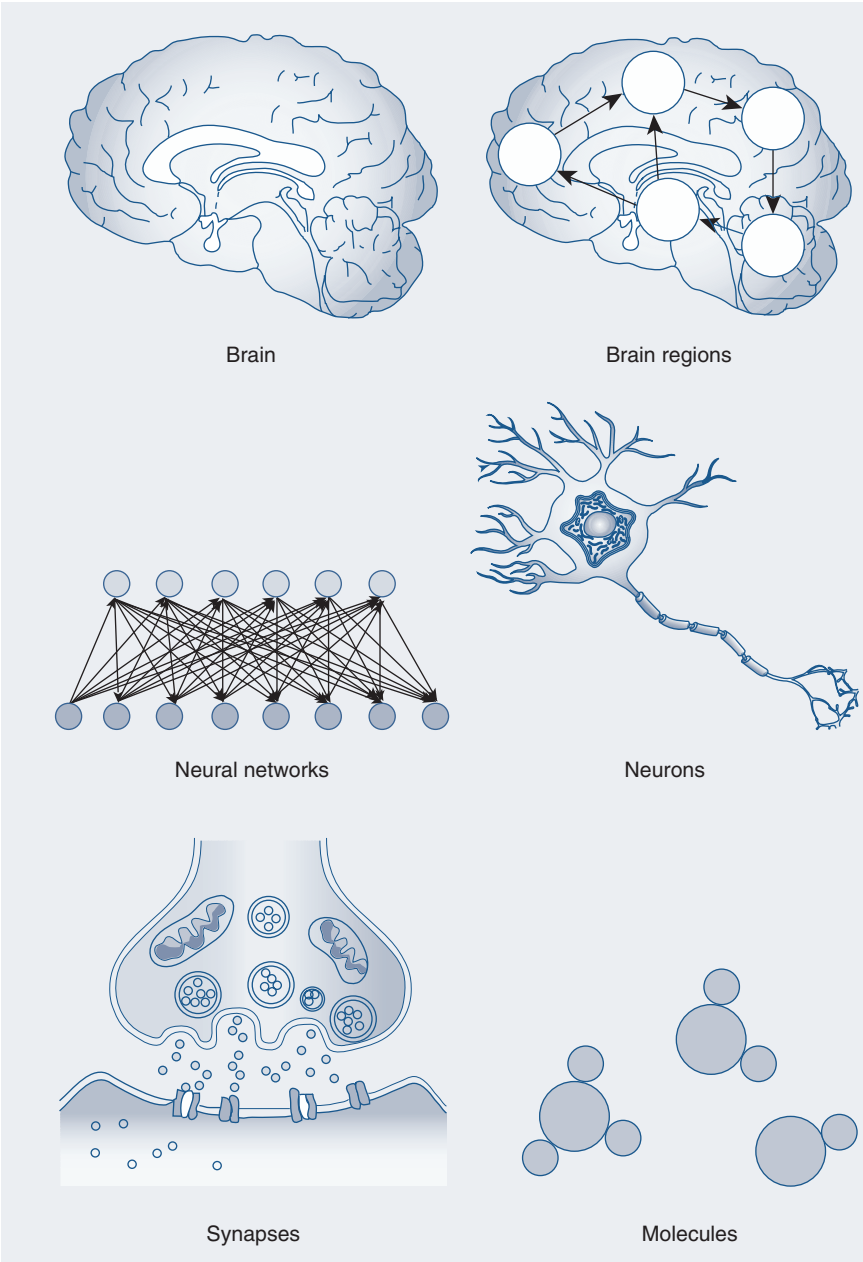
This partitioning of the analysis of information-processing events into three levels has been criticized as being fundamentally simplistic since the levels clearly interact with one another, and each level can, in turn, be further subdivided into its own sublevels (Churchland, Koch, & Sejnowski, 1990). Figure 1.3 depicts one possible organization of the many structural levels of analysis in the nervous system. Starting at the top, we might consider the brain as one organizational unit, brain regions as corresponding to another organizational unit one step down in spatial scale, then neural networks, individual neurons, and so on. Similarly, we could divide algorithmic steps into different substeps and problems into subproblems. To compound all this, it is not entirely clear how to map one level of analysis onto another. We may be able to specify clearly how an algorithm executes but be at a loss to say exactly where or how this is achieved with respect to the nervous system. David Marr's separation of the computational, algorithmic, and implementational levels can be a useful tool for studying cognition. But, like any tool, there may be parts of the task where it doesn't apply. (You don't want to find yourself using a hammer for a job that requires a screwdriver, or a laser scalpel.)

Differing Views of Representation and Computation

Before finishing our discussion of computation, it is important to differentiate between several different conceptions of what it is. So far, we have mostly been talking about computation as being based on the formal systems notion. In this view, a computer is a **formal symbol manipulator**. Let's break this definition down into its component parts. A system is formal if it is syntactic or rule governed. In general, we use the word **syntax** to refer to the set of rules that govern any symbol system. The rules of language and mathematics are formal systems because they specify which types of allowable changes can be made to symbols. Formal systems also operate on representations independent of the content of those representations. In other words, a process can be applied to a symbol regardless of its meaning or semantic content. A symbol, as we have already indicated, is a type of representation and can assume a wide variety of forms and can undergo a wide variety of manipulations. Manipulation here implies that computation is an active, physical process that takes place over time. That is, manipulations are actions, and they occur physically on some type of computing medium or substrate (e.g., neurons or circuits). And they take some time to occur (i.e., they don't happen instantaneously).

But this is not the only conception of what computation is. The connectionist or network approach to computation differs from the classical formal systems approach of cognitive science in several ways. In the classical view, knowledge is represented locally—in the form of symbols. In the connectionist view, knowledge is represented as a pattern of neural activation, or a pattern of synaptic strengths, that is distributed throughout a network and so is more global than a single symbol. Processing style is also different in this approach. The classical view has processing occurring in discrete sequential stages, whereas in connectionism, processing occurs in parallel through the simultaneous activation of many nodes or elements in the network. However, some cognitive scientists downplay these differences, arguing that information processing occurs in both systems

Figure 1.3 Structural levels of analysis in the nervous system.



and that the tri-level hypothesis can be applied equally to both (Dawson, 1998). We further compare and contrast the classical and connectionist views at the beginning of the network approach chapter.

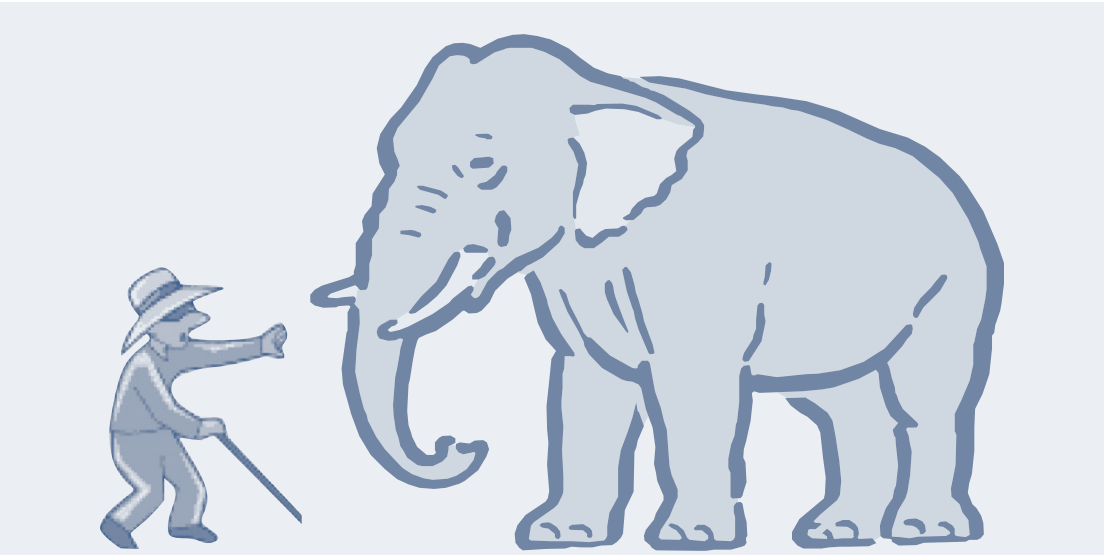
What happens to representations once they get established? In some versions of the symbolic and connectionist approaches, representations remain forever fixed and unchanging. For example, when you learned about the concept of a car, a particular symbol standing for "car" was formed. The physical form of this symbol could then be matched against perceptual inputs to allow you to recognize a car when you see one or used in thought processes to enable you to reason about cars when you need to. Such processes can be simplified if the symbol stays the same over time. This, in fact, is exactly how a computer deals with representation. Each of the numbers or letters of the alphabet has a unique and unchanging ASCII (American Standard Code for Information Interchange) code. Similarly, in artificial neural network simulations of mind, the networks are flexible and plastic during learning, but once learning is complete, they must remain the same. If not, the representations will be rewritten when new things are learned.

A third view of representation comes from the **dynamical perspective** in cognitive science (Casasanto & Lupyan, 2015; Friedenberg, 2009; Spivey, 2007). According to this view, the mind is constantly changing as it adapts to new information. A representation formed when we first learn a concept is altered each time we think about that concept or experience information that is in some way related to it. For example, let's say a child sees a car for the first time and it is red. The child may then think that all cars are red. The next time he or she sees a car, it is blue. The child's concept would then be changed to accommodate the new color. Even after we are very familiar with a concept, the context in which we experience it is always different. You may hear a friend discussing how fast one type of car is, or you may find yourself driving a car you have never driven before. In each of these cases, the internal representation is modified to account for the new experience. So it is unlikely that biological representations of the sort we see in humans will ever stay the same. We talk more about the dynamical systems approach and how it can unify differing perspectives on representation and computation in the ecological embodied chapter at the end of this book.

THE INTERDISCIPLINARY PERSPECTIVE

There is an old fable about five blind men who stumble on an elephant (see Figure 1.4). Not knowing what it is, they start to feel the animal. One man feels only the elephant's tusk and thinks he is feeling a giant carrot. A second man feels the ears and believes that the object is a big fan. The third feels the trunk and proclaims that it is a pestle, while a fourth touching only the leg believes that it is a mortar. The fifth man, touching the tail, has yet another opinion: He believes it to be a rope. Obviously, all five men are wrong in their conclusions because each has examined only one aspect of the elephant. If the five men had gotten together and shared their findings, they may easily have pieced together what kind of creature it was. This story serves as a nice metaphor for cognitive science. We can think of the elephant as the mind and the blind men as researchers in different disciplines in cognitive science. Each individual discipline may make great strides in understanding its particular subject matter but, if it cannot compare its results to those

Figure 1.4 If you were the blind man, would you know it is an elephant?



of other related disciplines, may miss out on understanding the real nature of what is being investigated.

The key, then, to figuring out something as mysterious and complex as mind is communication and cooperation among disciplines. This is what is meant when one talks about cognitive science—not the sum of each of the disciplines or approaches but, rather, their interaction. The best advances in our understanding have come from this kind of interdisciplinary cooperation. A number of major universities have established interdisciplinary cognitive science centers, where researchers in such diverse areas as philosophy, linguistics, neuroscience, computer science, and cognitive psychology are encouraged to work together on common problems. Each area can then contribute its unique strength to the phenomena under study. The consequent exchange of results and ideas then leads to fruitful synergies between these disciplines, accelerating progress with respect to finding solutions to the problem and yielding insights into other research questions.

We have alluded to some of the different approaches in cognitive science. Some of these approaches are longstanding disciplines in and of themselves, such as philosophy and psychology. But some of these approaches are perhaps better described as subdisciplines or research areas, such as the emotional approach to cognitive science and the embodied ecological approach to cognitive science. Because this book is about explaining each approach and its major theoretical contributions, it is worth mentioning each now in terms of its perspective, history, and methodology. In the following sections, we will also provide a brief preview of the issues addressed by each approach. In the rest of the book, we have devoted a chapter to each approach.

The Philosophical Approach

Philosophy is the oldest of all the disciplines in cognitive science. It traces its roots back to the ancient Greeks. Philosophers have been active throughout much of recorded human history, attempting to formulate and answer basic questions about the universe. This approach is free to study virtually any sort of important question on virtually any subject, ranging from the nature of existence to the acquisition of knowledge to politics, ethics, and beauty. Philosophers of mind narrow their focus to specific problems concerning the nature and characteristics of mind. They might ask questions such as, What is mind? How do we come to know things? How is mental knowledge organized?

The primary method of philosophical inquiry is reasoning, both deductive and inductive. **Deductive reasoning** involves the application of the rules of logic to statements about the world. Given an initial set of statements assumed to be true, philosophers can derive other statements that logically must be correct. For example, if the statement "College students study 3 hours every night" is true and the statement "Mary is a college student" is true, we can conclude that "Mary will study 3 hours every night." Philosophers also engage in **inductive reasoning**. They make observations about specific instances in the world, notice commonalities among them, and draw general conclusions. An example of inductive reasoning would be the following: "Whiskers the cat has four legs," "Scruffy the cat has four legs," and, therefore, "All cats have four legs." However, philosophers do not tend to use a systematic form of induction known as the scientific method. That is employed within the other cognitive science disciplines.

In Chapter 2, we summarize several of the fundamental issues facing philosophers of mind. With respect to the mind–body problem, philosophers wrangle over what exactly a mind is. Is the mind something physical like a rock or a chair, or is it nonphysical? Can minds exist only in brains, or can they emerge from the operation of other complex entities, such as computers? The knowledge acquisition problem deals with how we come to know things. Is knowledge a product of one's genetic endowment, or does it arise through one's interaction with the environment? How much does each of these factors contribute to any given mental ability? We also look into one of the most fascinating and enigmatic mysteries of mind—that of consciousness. In this case, we can ask, What is consciousness? Are we really conscious at all?

INTERDISCIPLINARY CROSSROADS: SCIENCE AND PHILOSOPHY

Philosophers have often been accused of doing "armchair" philosophy, meaning they do nothing but sit back in comfortable chairs and reflect about the world. The critique is that they ignore empirical evidence, information based on observation of the world, usually under

controlled circumstances as is the case in the sciences. This may be changing, however, with the development of a new trend known as experimental philosophy.

Experimental philosophy uses empirical methods, typically in the form of surveys that assess people's understanding of constructed scenarios, to help answer philosophical questions. Followers of the movement call themselves "X-Philes" and have even adopted a burning armchair and accompanying song as their motto, a video of which is viewable on YouTube.

One of the more prominent X-Philers, Joshua Knobe (2003), presented participants with a hypothetical situation. In the situation, a vice president of a company approaches the chairman of the board and asks about starting a new program. The program will increase profits, but it will also harm the environment. In one version of the survey, the chairman responds that he or she doesn't care at all about harming the environment but just wants to make as much profit as possible and so wants to begin the new program.

When participants were asked whether the chairman harmed the environment intentionally, 82% responded "yes." In a different version of the scenario, everything was identical except that "harming" was now replaced with "helping." The program had the side effect of helping the environment but the chairman again stated that he or she didn't care at all about helping the environment and wanted to go ahead with it. This time when asked if the chairman intentionally helped the environment, only 23% responded "yes."

Logically, there should be no difference in perceived intentionality between both conditions. In each case, there was a desire to make profit and a failure to care one way or the other about the environment. But clearly, the way we see intentionality involves moral evaluations, the way people judge what is good or bad. This conclusion, what is now called the "Knobe effect," could not necessarily have been determined a priori using pure reasoning. The term *a priori* relates to what can be known through an understanding of how certain things work rather than by observation.

Experimental philosophy has addressed a wide variety of issues. These include cultural differences in how people know or merely believe (Weinberg, Nichols, & Stich, 2001), whether a person can be morally responsible if his or her actions are determined (Nichols & Knobe, 2007), the ways in which people understand consciousness (Huebner, 2010), and even how philosophical intuitions are related to personality traits (Feltz & Cokely, 2009).

Of course, this new field has been critiqued. Much of experimental philosophy has merely used questionnaires that tap into the intuitions of experimental participants (much like social psychology already does). Williamson (2008) argues that philosophical evidence should not just rely on people's intuitions. Another obvious issue is that experimental philosophy is not philosophy at all, but science. In other words, if it is adopting the use of survey methodology, statistics, and so on, then it really becomes psychology or a branch of the social sciences rather than philosophy.

The Psychological Approach

Compared with philosophy, psychology is a relatively young discipline. It can be considered old, though, particularly when compared with some of the more recent newcomers to the cognitive science scene—for example, AI and robotics. Psychology as a science

arose in the late 19th century and was the first discipline in which the scientific method was applied exclusively to the study of mental phenomena. Early psychologists established experimental laboratories that would enable them to catalog mental ideas and investigate various mental capacities, such as vision and memory. Psychologists apply the scientific method to both mind and behavior. That is, they attempt to understand not just internal mental phenomena, such as thoughts, but also the external behaviors that these internal phenomena can give rise to.

The scientific method is a way of getting hold of valid knowledge about the world. One starts with a hypothesis or idea about how the world works and then designs an experiment to see if the hypothesis has validity. In an experiment, one essentially makes observations under a set of controlled conditions. The resulting data then either support or fail to support the hypothesis. This procedure, employed within psychology, and cognitive science in general, is described more fully at the start of Chapter 3.

The field of psychology is broad and encompasses many subdisciplines, each one having its unique theoretical orientations. Each discipline has a different take on what mind is. The earliest psychologists—that is, the voluntarists and structuralists—viewed the mind as a kind of test tube in which chemical reactions between mental elements took place. In contrast, functionalism viewed mind not according to its constituent parts but, rather, according to what its operations were—what it could do. The Gestaltists again went back to a vision of mind as composed of parts but emphasized that it was the combination and interaction of the parts, which gave rise to new wholes, that was important. Psychoanalytic psychology conceives of mind as a collection of differing and competing mind-like subsystems, while behaviorism sees it as something that maps stimuli onto responses.

The Cognitive Approach

Starting in the 1960s, a new form of psychology arrived on the scene. Known as cognitive psychology, it came into being, in part, as a backlash against the behaviorist movement and its profound emphasis on behavior (and its neglect of mental computations). Cognitive psychologists placed renewed emphasis on the study of internal mental operations. They adopted the computer as a metaphor for mind and described mental functioning in terms of representation and computation. They believed that the mind, like a computer, could be understood in terms of information processing.

The cognitive approach was also better able to explain phenomena such as language acquisition for which behaviorists did not have good accounts. At around the same time, new technologies that allowed better measurement of mental activity were being developed. This promoted a movement away from the behaviorist's emphasis solely on external observable responses and toward the cognitive scientist's emphasis on internal functions, as these could, for the first time, be observed with reasonable precision.

Inherent early on in the cognitive approach was the idea of modularity. Modules are functionally independent units that receive inputs from other modules, perform a specific processing task, and pass the results of their computation onto other modules. The influence of the modular approach can be seen in the use of process models or flow

diagrams. These depict a given mental activity via the use of boxes and arrows, where boxes depict modules and arrows depict the flow of information among them. The techniques used in this approach are the experimental method and computational modeling. Computational modeling involves carrying out a formal (typically software-based) implementation of a proposed cognitive process. Researchers can run the modeling process so as to simulate how the process might operate in a human mind. They can then alter various parameters of the model or change its structure in an effort to achieve results as close as possible to those obtained in human experiments. The model can also produce new results that then inspire new experiments. This use of modeling and comparison with experimental data is a key strength in cognitive psychology and is also used in the AI and artificial network approaches.

Cognitive psychologists have studied a wide variety of mental processes, including pattern recognition, attention, memory, imagery, and problem solving. Theoretical accounts and processing models for each of these are given in Chapters 4 and 5. Language is also within the purview of cognitive psychology, but because the approach to language is so multidisciplinary, we describe it separately in Chapter 9.

The Neuroscience Approach

Brain anatomy and physiology have been studied for centuries. It is only recently, however, that we have seen tremendous advances in our understanding of the brain, especially in terms of how neuronal processes can account for cognitive phenomena. The study of the brain and endocrine system and how these account for mental states and behavior is called neuroscience. The attempt to explain cognitive processes in terms of underlying brain mechanisms is known as cognitive neuroscience.

Neuroscience, first and foremost, provides a description of mental events at the implementational level. It attempts to describe the biological "hardware" on which mental "software" supposedly runs. However, as discussed above, there are many levels of scale when it comes to describing the brain, and it is not always clear which level provides the best explanation for any given cognitive process. Neuroscientists, however, investigate at each of these levels. They study the chemistry of neurotransmitters, the cell biology of individual neurons, the process of neuron-to-neuron synaptic transmission, the patterns of activity in local cell populations, and the interrelations of larger brain areas.

A reason for many of the recent developments in neuroscience is, again, the development of new technologies. Neuroscientists employ a wide variety of methods to measure the performance of the brain at work. These include electroencephalography (EEG) that detects changes in the electric fields generated by large groups of highly active neurons, magnetoencephalography (MEG) that detects changes in the magnetic fields generated by groups of highly active neurons, and functional magnetic resonance imaging (fMRI) that detects changes in the magnetic properties of oxygenated blood being delivered to small groups of highly active neurons, as well as many others. Studies that use these procedures have participants perform a cognitive task, and the brain activity that is concurrent with the performance of the task is recorded. For example, a participant may be asked to form a mental image of a particular object with their

eyes closed. The researchers can then determine which parts of the brain become active during visual imagery and in what order. Neuroscientists use other techniques as well. They study brain-damaged patients and the effects of lesions in laboratory animals, and they use single- and multiple-cell recording techniques.

The Network Approach

The network approach is at least partially derived from neuroscience. In this perspective, mind is seen as a collection of computing units. These units are connected to one another and mutually influence one another's activity via their connections, although each of the units is believed to perform a relatively simple computation—for example, a neuron can either "fire" by initiating an action potential that sends an electrochemical signal to other neurons or not "fire" and fail to initiate an action potential. In these networks, the connectivity among many units can give rise to representational and computational complexity.

Chapter 7, which outlines the network approach, has two parts. The first involves the construction of artificial neural networks. Most artificial neural networks are computer software simulations that have been designed to mimic the way actual brain networks operate. They attempt to simulate the functioning of neural cell populations. Artificial neural networks that can perform arithmetic, recognize faces, learn concepts, play competitive backgammon, and read out loud now exist. A wide variety of network architectures has developed over the past 30 years.

The second part of the network chapter is more theoretical and focuses on knowledge representation—on how meaningful information may be mentally encoded and processed. In semantic networks, nodes standing for concepts are connected to one another in such a way that activation of one node causes activation of other related nodes. Semantic networks have been constructed to explain how conceptual information in memory is organized and recalled. They are often used to predict and explain data obtained from experiments with human participants in cognitive psychology.

In the chapter on networks, we will finish off with a discussion of network science. This is a new interdisciplinary field, much like cognitive science itself. However, in this field, researchers focus on the structure and function of networks. The term *network* is meant in a very broad sense here to include not just artificial or natural *neural* networks but also telephone and wireless networks, electrical power networks, ecosystem networks, and human social networks. Surprisingly, we will see that there are commonalities among these different types of networks and that they share some organizational and operational features. We will examine these features and apply them particularly to the brain and cognition.

The Evolutionary Approach

The theory of natural selection proposed by Charles Darwin in 1859 revolutionized our way of thinking about biology. Natural selection holds that adaptive features enable the animals that possess them to survive and pass these features on to future generations. The environment, in this view, is seen as selecting from among a variety of traits those that serve a functional purpose.

The field of evolutionary psychology applies the theory of natural selection to account for human mental processes. It attempts to elucidate the selection forces that acted on our ancestors and how those forces gave rise to the cognitive structures we now possess. Evolutionary psychologists tend to adopt a modular approach to mind. In this case, the proposed modules would be individual mechanisms in the brain that carry out various distinct cognitive abilities that were successful at solving certain problems, thus helping our ancestors contribute their genes to the next generation. Evolutionary theories have been proposed to account for experimental results across a wide range of capacities, from categorization to memory to logical and probabilistic reasoning, language, and cognitive differences between the sexes. They also have been proposed to account for how we reason about money and other resources—a new field of study known as behavioral economics.

Also in this chapter, we examine comparative cognition. This is the study of animal intelligence. We look at the cognitive capacities of a number of different species and discuss some of the problems that arise in comparing animals with one another and with humans.

The Linguistic Approach

Linguistics is an area that focuses exclusively on the domain of language. It is concerned with all questions concerning language ability, such as, What is language made of? How do we acquire language? What parts of the brain underlie language use? As we have seen, language is also a topic studied within other disciplines—for example, cognitive psychology and neuroscience. Because so many different researchers in different disciplines have taken on the problem of language, we consider it here as a separate discipline, united more by topic than by perspective or methodology.

Part of the difficulty in studying language is the fact that language itself is so complex. Much research has been devoted to understanding its nature. This work looks at the properties all languages share, the elements of language, and how those elements are used during communication. Other foci of linguistic investigation center on language acquisition, deficits in language acquisition caused by early sensory deprivation or brain damage, the relationship between language and thought, language use by nonhuman primates, and the development of automated speech recognition systems.

Linguistics, perhaps more than any other perspective discussed here, adopts a very eclectic interdisciplinary methodological approach. Language researchers employ experiments and computer models, study brain-damaged patients, track how language ability changes during development, and compare diverse languages.

The Emotion Approach

As you may have surmised, humans don't just think—we also feel. Our conscious experience consists of emotions, such as happiness, sadness, and anger. Recent work in cognitive psychology and other fields has produced a wealth of data on emotions and how they influence thoughts. In Chapter 10, we start out by examining what emotions are and how they differ from moods. We examine several different theories of emotion and describe

how they influence perception, attention, memory, and decision making. Following this, we look at the neuroscience underlying emotions and the role that evolutionary forces played in their formation. AI investigators have formulated models of how computers can "compute" and display emotional behavior. There are even robots capable of interacting with people in an emotionally interactive manner.

The Social Approach

Much of cognition happens inside individuals. But if some of it is happening in between an individual and his or her environment, then the cognitive relations *between* individuals clearly constitute an important aspect of cognition. The field of social cognition explores how people make sense of both themselves and others. We will see that thinking about people often differs from thinking about objects and that different parts of our brains are used when thinking socially. Neuroscience has revealed that in laboratory animals, specialized cells, called mirror neurons, are active both when an animal performs some action and when it watches another animal perform that same action. Later in Chapter 11, we introduce the concept of a theory of mind. This is an ability to understand and appreciate other people's states of mind. This capacity may be lacking in people suffering from autism. We conclude by summarizing work on specific social cognitive phenomena: attitudes, impressions, attributions, stereotypes, and prejudice.

The Artificial Intelligence Approach

Researchers have been building devices that attempt to mimic human and animal function for many centuries. But it is only in the past few decades that computer scientists have seriously attempted to build devices that mimic complex thought processes. This area is now known as artificial intelligence. Researchers in AI are concerned with getting computers to perform tasks that have heretofore required human intelligence. As such, they construct programs to do the sorts of things that require complex reasoning on our part. AI programs have been developed that can diagnose medical disorders, use language, and play chess.

AI also gives us insights into the function of human mental operations. Designing a computer program that can visually recognize an object often proves useful in understanding how we may perform the same task ourselves. An even more exciting outcome of AI research is that someday, we may be able to create an artificial person who will possess all or many of the features that we consider uniquely human, such as consciousness, the ability to make decisions, and so on (Friedenberg, 2008).

The methods employed in the AI perspective include the development and testing of computer algorithms, their comparison with empirical data or performance standards, and their subsequent modification. Researchers have employed a wide range of approaches. An early attempt at getting computers to reason involved the application of logical rules to propositional statements. Later, other techniques were used. Chapters 12 and 13 give detailed descriptions of these techniques.

The Robotics Approach

Finally, we consider robotics. Robotics may be considered a familial relation to AI and has appeared on the scene as a formal discipline only recently. Whereas AI workers build devices that "think," robotics researchers build machines that must also "act." In fact, the embodied ecological approach has assisted roboticists in discovering that adding a robot body to an AI system can actually help the AI system think better and learn faster. Investigators in this field build autonomous or semiautonomous mechanical devices that have been designed to perform a physical task in a real-world environment. Examples of things that robots can do presently include walking through obstacle courses, competing in miniature soccer games, driving through city streets, welding or manipulating parts on an assembly line, performing search and rescue in hazardous environments, defusing bombs, vacuuming your apartment, and destroying each other on television.

The robotics approach has much to contribute to cognitive science and to theories of mind. Robots, like people and animals, must demonstrate successful goal-oriented behaviors under complex, changing, and uncertain environmental conditions. Robotics, therefore, helps us think about the kinds of minds that underlie and produce such behaviors.

In Chapter 13, we outline different paradigms in robotics. Some of these approaches differ radically from one another. The **hierarchical paradigm** offers a "top-down" perspective, according to which a robot is programmed with knowledge about the world. The robot then uses this model or internal representation to guide its actions. The **reactive paradigm**, on the other hand, is "bottom up." Robots that use this architecture respond in a simple way to environmental stimuli: They react reflexively to a stimulus input, and there is little in the way of intervening knowledge.

The Embodied Ecological Approach

Inspired in part by the reactive paradigm in robotics, cognitive scientists have been discovering that the body itself plays an important role in cognition. In the embodied cognition approach, your body does some of your thinking for you, not just your brain. The particular ways in which the body interfaces with the environment are partly responsible for the cognition that emerges. For example, while part of how a person recognizes an apple may involve mental representations of its visual features (e.g., curved contours and reddish hue), another part of how that apple is recognized is by partially activating the actions that the hand and mouth would typically perform on it. So if you had a different body (e.g., different height, different strength, different number of limbs, different kind of eyes), the way you think would be a little different. In fact, even a tool in your well-trained hand can be treated as part of who you are. These insights about non-brain-based cognitive processes are couched well in the context of a perspective on perception-and-action called the ecological approach. In the ecological approach, intelligent behavior emerges not only from information inside the brain, and inside the body, but also from information inherent in the *relationship* between the body and the environment. For instance, when a person is standing with their eyes open, light reflects off of surfaces around them and projects a pattern of information on their retinas. When

that person begins walking, that pattern of retinal information changes in a systematic fashion due specifically to the way the person is walking. Rather than passively receiving that sensory input like a computational information processor, the active observer is in charge of how that information changes over time. If she chooses to walk fast, the retinal information changes quickly. If she chooses to walk slow, the retinal information changes slowly. If she chooses to walk with a bounce, the retinal information changes differently still. These insights on cognition are obtained only by focusing on the organism's relationship to its ecological environment, not by focusing on the organism in isolation. And there's more than just apples, tools, and sidewalks in your ecological environment that become part of your cognition. There's other people too. In Chapter 14, the embodied ecological approach is shown to draw its insights from dynamical systems theory, philosophy, psychology, neuroscience, linguistics, and even anthropology.

Integrating Approaches

Many of the approaches we have just listed inform one another. For instance, the fields of AI and robotics are in some cases inseparable. AI programmers often write computer programs that serve as the "brains" of robots, telling them what to do and providing them with instructions on how to perform various tasks like object recognition and manipulation. In recent years, the cognitive and neuroscience approaches have come closer together, with cognitive psychologists providing information-processing models of specific brain processes. For example, there are cognitive models of hippocampal function that specify how this brain region helps the brain encode memories based on our understanding of the neural substrates. In an even more interdisciplinary leap, there is the new area of social cognitive neuroscience in which social contexts are attached to cognitive–neural models. At the end of Chapter 14, we provide some insights into the ways in which these different approaches are and can be integrated. The integration of these different disciplines and approaches may point toward new ways to understand how cognition works.

SUMMING UP: A REVIEW OF CHAPTER 1

1. Cognitive science is the scientific interdisciplinary study of mind and sees contributions from multiple fields, including philosophy (along with its newest offshoot, experimental philosophy), psychology, cognitive psychology, neuroscience, the connectionist or network approach, evolution, linguistics, embodied cognition and ecological perception, the scientific study of emotions and social behavior, AI, robotics, and more.
2. Mind can be considered an information processor. At least some mental operations bear some similarity to the way information is processed in a computer.

3. Information processing requires that some aspect of the world be represented and then operated on or computed. A representation is symbolic if it stands for something else. The thing a symbol stands for in the world is called its referent. The fact that symbols are "about" these things is called intentionality.
4. A formal system is made of symbols—collections of symbols that form expressions and processes that act on those expressions to form new expressions. Formal logic is an example of a formal system.
5. According to the physical symbol system hypothesis, formal systems can be said to be intelligent, implying that computers may be intelligent. A problem for this is the symbol grounding problem, which states that symbols in people are grounded because we have bodies, can perceive objects, and can act on them. Therefore, if a computer were to be embodied with sensors and effectors that allow it to actively gather information from the environment, then it too may be able to ground its symbols.
6. Examples of representations are concepts, propositions, rules, and analogies. Representations are realized by an information bearer, have content, are grounded, and need to be interpreted.
7. Computations are processes that act on or transform representations. According to the tri-level hypothesis, there may be at least three levels of analysis for information processing systems: computational, algorithmic, and implementational.
8. Several different schools of thought differ in the way they view representation and computation. In the classical cognitive science view, representations are fixed symbols and information processing is serial. In the connectionist view, representations are distributed and processing is parallel. According to the dynamical view, representations are constantly changing, being altered with each new experience.

SUGGESTED READINGS

Friedenberg, J. (2009). *Dynamical psychology: Complexity, self-organization, and mind.* Charlotte, NC: Emergent.

Harnish, R. M. (2002). *Minds, brains, computers: An introduction to the foundations of cognitive science.* Malden, MA: Blackwell.

Sobel, C. P. (2001). *The cognitive sciences: An interdisciplinary approach.* Mountain View, CA: Mayfield.

Stainton, R. J. (2006). *Contemporary debates in cognitive science.* Malden, MA: Blackwell.

CHAPTER TWO

THE PHILOSOPHICAL APPROACH

Enduring Questions

WHAT IS PHILOSOPHY?

Philosophy in its broadest sense is the search for wisdom and knowledge. It is the first approach we will tackle in our voyage through the different disciplines of cognitive science. There are good reasons for beginning here. Philosophy plays a vital participatory role in cognitive science. It does this not by generating results, since it is a theoretical rather than experimental discipline, but by "defining problems, criticizing models, and suggesting areas for future research" (Garfield, 1995, p. 374). More than any other discipline in cognitive science, philosophy is not limited by its subject matter or a particular theoretical stance. It is, therefore, free to evaluate and contribute to the remaining disciplines in a way the others cannot. This approach is also the oldest of the different approaches, tracing its origins back to the ancient Greeks. It is thus fitting that we begin our tour here.

The translation of the word *philosophy* yields "love of wisdom," indicating the philosopher's concern with knowledge and understanding the universe. Philosophy as a formal discipline studies a wide range of topics. In fact, there is no topic that is not fair game for a philosopher; he or she may examine politics, ethics, esthetics, and other subjects. We concern ourselves here with two branches of philosophy. **Metaphysics** examines the nature of reality. The mind–body problem is a metaphysical one at heart because it seeks to understand whether the mental world is part of the physical material world. **Epistemology** is the study of knowledge and asks questions such as, What is knowledge? How is knowledge represented in the mind? How do we come to acquire knowledge?

CHAPTER OVERVIEW

In this chapter, we survey philosophic thoughts that center on a couple of vexing issues, which can be summed up in terms of "this"

Learning Objectives

After reading this chapter, you will be able to:

1. Define the mind–body problem.
2. Compare monism and dualism as potential solutions to the mind–body problem.
3. Assess functionalism as a framework for studying the mind.
4. Formulate an example syllogism that is logically sound.
5. Describe the "hard problem" of consciousness.
6. Contrast reductionism and emergence.
7. Describe and critique Searle's Chinese room thought experiment.

versus "that." Such terminology suggests that the debates that have arisen from these issues have polarized the arguments and that there are only two possible answers to a problem. We will see that this is actually not the case and that there are multiple ways to conceptualize the issues. These issues are the mind–body and nature–nurture debates. In addition, we discuss the question of consciousness and its relation to cognitive science.

This chapter is motivated by great questions in the philosophy of mind. In the first section on the mind–body problem, we ask a very fundamental question: What is mind? Is it something that is physical or not? Is a body necessary to have a mind? These questions are primarily metaphysical in nature. In the second section on functionalism, we get more specific. If a mind is physical, then what sort of substrate can it be based on? Can a mind only be based on brains or can minds emerge from other things like computers? In the third section, we address the issue of knowledge. How does information get into our "heads"? Are we born knowing certain things or is information primarily learned? These are epistemological questions, since they concern information. In the fourth and final section, we look at perhaps the most contentious and debated issue in the modern philosophy of mind—consciousness. Here we will address questions like What is consciousness? Can it exist in other animals? What exactly is happening in our brains when we have a particular conscious state?

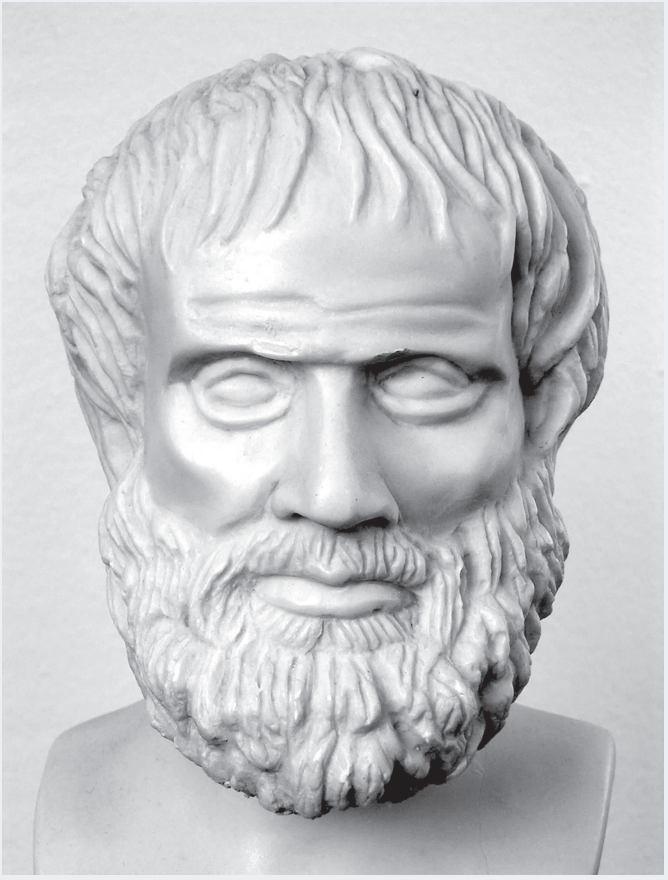
THE MIND–BODY PROBLEM: WHAT IS MIND?

The mind–body problem addresses how psychological or mental properties are related to physical properties. The debate stems from a fundamental conception about what the mind is. On the one hand, we have the brain that is material and physical. It is made up of substances that we can measure and understand. The mind could be thought of in the same way, as simply a physical thing. On the other hand, there are those who argue that the mind is something more. They say we can't equate our subjective conscious experiences, such as beliefs, desires, and thoughts, with something as mundane as the human nervous system. They say that the mind is nonphysical and consists of something resembling a soul or spirit. The mind as a nonphysical entity inhabiting the brain or some other physical system is sometimes called "the ghost in the machine."

The first question of the mind–body problem is metaphysical and refers to the nature of what mind is. Is the mind physical or something else? A second and more specific question concerns the relationship between these two things. If we were to assume that there are two such entities, then what is the causal relationship between them? Does the mind control the body, or does the body control the mind? There are many theories supporting the directions this control takes. Some theories argue that only one exerts control; others state that they both control each other; additional theories state that the two work in parallel but that neither has any causal influence.

Our discussion in this section is structured around basic conceptions of the nature of mind. According to **monism**, there is only one kind of state or substance in the universe. The ancient Greek philosopher Aristotle (384–322 BCE) was a monist (see Figure 2.1). He characterized the difference between mind and body as the difference between form and matter. One way to think of his notion is to consider a lump of clay. It

Figure 2.1 **A bust of the early Greek philosopher Aristotle. He believed that there was no substantial difference between mind and matter.**



is made up of physical matter, and we can think of it as corresponding to the brain. We can shape the clay with our hands into different forms—for example, we can roll it into a ball or flatten it into a pancake. The shapes the clay can assume, Aristotle implied, are like the different thoughts the mind can take on when it undergoes different patterns of activity. These shapes are just different physical states and do not constitute any non-physical or spiritual substance.

In **dualism**, one believes that both mental and physical substances are possible. Plato, another Greek philosopher (427–347 BCE), was a dualist. Plato was Aristotle's

teacher, but the two held quite different views. Plato believed that the mind and the body exist in two separate worlds. Knowledge of the mind, he thought, exists in an ideal world of forms, which is immaterial, nonextended, and eternal. The body instead resides in a world that is material, extended, and perishable. There are crucial differences between the objects of one world and those of the other. Mental ideas such as "circle" that reside in the ideal world of forms are perfect, and according to Plato, the circles of this world are always perfectly round. Concrete examples of circles that we find in the real world are always imperfect. If we examine an actual circle, at some level of magnification, the circle's edge will lose its curvature.

A possible way out of the mind–body problem is to rely solely on behaviors as indicators of mental states. Behaviors such as talking, running, and laughing are external and more easily measured than brain activity. According to **philosophical behaviorism**, mental states are dispositions or tendencies to behave in certain ways under certain circumstances. Happiness, for instance, can be thought of as the tendency to smile or laugh. There is support for this idea in that there are strong connections between mental states and behavior. Many people do smile when they are happy or cry when they are sad. However, we cannot conclusively prove any causal link between the two. A "tough guy" or stoic might not cry when sad for fear of being perceived as weak. Alternatively, somebody might exhibit nervous laughter before going on stage, in which case anxiety and not happiness is causing the laughter. For these reasons, we cannot consistently rely on behaviors as indicators of mental states. Please note that philosophical behaviorism is not the same as psychological behaviorism, discussed in Chapter 3. The psychological version focuses on the laws that govern the relationship between stimuli and responses and how we can shape responses by manipulating stimuli such as rewards and punishments. Psychological behaviorism ignores mental states altogether.

MONISM

If we are monists, we are left with two fundamental choices. Either the universe is mental or it is physical. It is difficult to take the first argument, called **idealism**, seriously—although it cannot be falsified. Imagine that everything you know isn't real but is simply an illusion of reality. The world you perceive and understand exists only "in your head"—although, in this universe, you don't have a real head, so we can say this only colloquially. A supreme being such as God could be responsible for this. Alternately, it could be that aliens have taken our brains and are feeding them information in such a way as to make us think that we live in the world we do. This is known informally as the "brain in a vat argument." In either case, we could never get outside our own conscious experience to prove what reality is really like.

On the flip side of our metaphysical coin is **physicalism**, or materialism. The origins of this view go back to the Greek philosopher Democritus (ca. 460–370 BCE), who believed that all things were composed of atoms. The attributes and behaviors of the atoms, he said, can explain the differences between things. Physicalists are also monistic and believe that the universe is composed of a single substance. However, they regard this substance as physical and material rather than spiritual or ethereal. Physicalism is,

thus, the doctrine that everything that exists is physical. The operations of the human mind are seen here simply as the operations of the human nervous system.

We can consider identity theory to be a form of physicalism. According to **identity theory**, the mind *is* the brain. Specifically, mental states are physical states of the brain. Just as salt is the same as its chemical structure of sodium chloride, in this perspective, we say that being happy is identical to a particular type of brain activity. In fact, for every type of what we think are mental states, there is a type of brain state. These physical brain states cannot be discovered through simple observation or an examination of the linguistic meaning of the terms. They can be understood only through proper scientific analysis.

Anyone who believes in identity theory will be left with more terminology than is needed. On the one hand, we have the common everyday subjective terms we all use to refer to mental states. These are words such as "feeling good" that we use to denote our subjective experiences. On the other hand, we have objective scientific phrases such as "activation of dopaminergic neurons in the meso-cortico-limbic system" to describe the brain states for these terms. Because the scientific descriptions are more accurate and better reflect what is actually going on, it might make sense to abandon our mental terms and rely solely on the scientific terminology. In effect, this is what **eliminativism** is about. Eliminativists reject not only the terminology but also the mental states they supposedly describe. According to eliminativism, once you have a good biological explanation for a psychological phenomenon, then you can eliminate the psychological explanation and rely solely on the biological one. Eliminativists don't believe in mental states at all. Only physical brain states are acknowledged as existing.

Theories of mind that use subjective terms such as those mentioned above and that use commonsense or intuitive reasoning are collectively referred to as **folk psychology**. Here is an example of a folk psychology theory: Cookies taste good. Robert likes to feel good. If Robert walks by the cookie store on his way home from school, he will buy and then eat some cookies. Folk psychology helps us understand and predict the behavior of people around us. In this sense, it is very useful. However, folk psychology is far removed from rigorous scientific explanations that are tested using experimental methods.

Churchland (1981) argues that folk psychology has outlived its usefulness and that we should eliminate it and replace it with neuroscience. He argues that folk psychology has not yielded any new insights into human behavior. It fails to inform us on a wide range of issues such as mental illness, creativity, and memory. Unlike valid scientific theories that tend to support one another, no folk psychology theories support one another or support scientific theories. Folk psychology, however, may still prove its worth in that it may serve as a basis for generating scientifically testable hypotheses. This would work only if the translation between the ill-defined subjective notions could be translated effectively into well-defined hypotheses that can be subjected to scientific testing.

Evaluating the Monist Perspective

Monist views have one clear advantage: They are simpler than dualist accounts. Rather than partitioning the universe into two types of things, they allow for only a single self-consistent universe. This follows the principle of Occam's razor, which states that all things being equal, the simplest explanation is usually the correct one. As mentioned

above, idealism does not really qualify as a legitimate theory because it cannot be proved true or false. As a result, it doesn't serve as a legitimate subject of scientific scrutiny. On the other hand, there is abundant evidence in favor of physicalism. Indeed, most of the experimental results in this book support the notion of the brain as being the core engine of mind. For example, neurological studies that look at behavior following brain damage show that when particular parts of the brain are damaged, relatively specific types of cognitive deficits can follow. This suggests that those brain areas are important physical mechanisms for those particular mental phenomena.

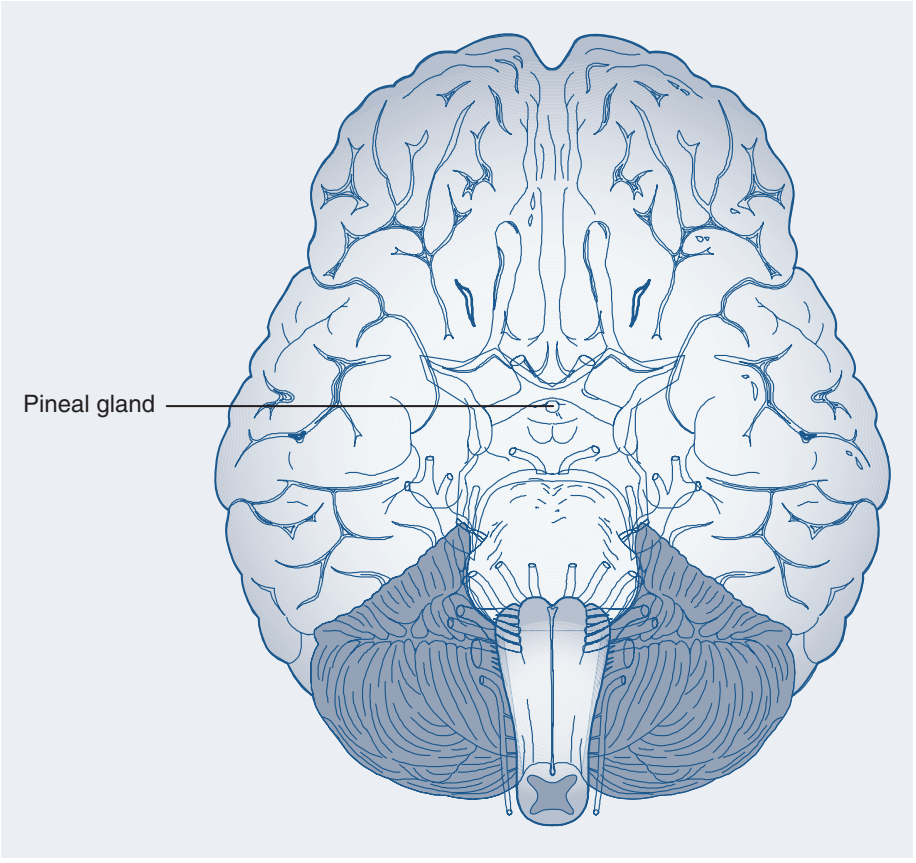
Physicalism, however, also has its share of critics. Some allow that physical processes can determine mental ones but deny that they can explain them. So, they might argue, changes in the brain may very well correlate with and produce fear, but they do not explain different kinds of fear, how a person becomes fearful, and so on. These critics acknowledge that the world is physical but indicate that there is no physical explanation for many phenomena. In these cases, they believe, it is perhaps better to explain using mental terms.

One argument that can be directed at identity theory goes by the name of **multiple realization**. Simply put, any given mental state, such as pain, can be instantiated or realized differently in different creatures. Pain in humans may be identical to activity in c-fibers, pain in squid may be identical to activity in d-fibers, while pain in aliens may be identical to activity in e-fibers (Ravenscroft, 2005). Furthermore, it is quite likely that differences exist not only between species but also between individuals within a given species. You and I may both feel pain when there is activity in c-fibers, but the particular pattern of activity in those fibers and in other areas of the brain could produce radically different mental perceptions of pain in each of us. Identity theory must, therefore, account for all the subtle mental differences that can exist between individuals.

DUALISM

Now that we have reviewed the various forms of monism, let us turn our attention to its logical alternative. Dualists believe both mental and physical realms are possible but differ in the way they think these two interact. **Classical dualism** originated with the French philosopher René Descartes (1596–1650). Descartes was a revolutionary philosopher for his time and introduced theories on many of the ideas that underlie cognitive science. He believed in a one-way causal link, with the nonphysical mind controlling the physical body but not vice versa. Descartes thought the mind exerted its control on the body through the pineal gland, perhaps because it is one of the few anatomical structures not duplicated on either side of the brain (see Figure 2.2). In this view, the mind is like a puppet master, the body is like a puppet, and the pineal gland is like the puppet strings by which the former controls the latter. Classical dualism conforms to most people's commonsense notion of the mind–brain relationship, which is that our thoughts control our actions. For instance, when we feel hungry, we get up and eat a snack. It certainly seems as though the feeling of hunger comes first and causes the action of getting up to eat.

Figure 2.2 **Descartes believed the pineal gland to be the location where the mind influenced the body. This belief was itself influenced by the fact that the pineal gland is located in the center of the brain.**



Substance Dualism

Descartes's views conform to what we now more commonly call **substance dualism**, where mind and body are composed of completely different substances. By *substance*, we mean the actual "stuff" that makes them up. In this view, there are two sorts of stuff: mental substances and physical substances. The physical substances can be the atoms, molecules, and cells of the natural world as we know it. It isn't clear what the basic elements or building blocks of mental substance might be.

There is a logical argument in favor of substance dualism that takes the form of several propositions. It goes as follows: Minds can do X. No physical object can do X. Therefore, minds are not physical objects. X can be any mental ability that has not been fully realized in some alternate physical form. Candidates include language, reason, and consciousness. The problem with this argument is that many of the abilities that we thought were exclusively human have now been replicated in machine form. Artificial intelligence (AI) is now capable of pattern recognition, language comprehension and production, and even artistic creativity (Friedenberg, 2008).

The arguments against substance dualism are stronger. One early attack came from Princess Elizabeth of Bohemia (1618–1680). She pointed out to Descartes that if mind and body are of two different substances, then they should not be able to causally interact. Saying they interact at the pineal gland specifies the location at which this might occur but not how. Descartes was unable to come up with a successful reply.

At least two other arguments against dualism involve explanatory power. First, biology can give us a very good account of mental states. If we can explain these states completely in terms of neurons, action potentials, and synapses, then there is no need to even provide a nonphysical account. Second, substance dualism does not give us any kind of alternate explanation. What are mental things made of? How do these things interact or operate? These questions are left unanswered.

Property Dualism

Another type of dualism may overcome these difficulties. According to **property dualism**, the mind and the body can be of the same stuff but have different properties. By *properties*, we mean characteristics or features. To illustrate, a golf ball and a tennis ball are both made of atoms, yet they differ in their properties. A golf ball is smaller, white, and dimpled. A tennis ball is larger, green, and fuzzy.

A property dualist believes that mental states are nonphysical properties of the brain. He or she would say that the brain has certain properties, such as being gray and wrinkled, weighing a certain number of pounds, and so on. But then he or she would also say that the brain gives rise to another set of properties, such as seeing red, being in pain, and feeling happy, that are wholly different from any of the currently understood properties of the brain.

Very little is gained in the dualist switch from substance to property. Nonphysicality still fails to provide an explanation for mental states, whether those are characterized as features or not. Again, we can ask, How do physical brain processes give rise to mental features? We are still left with better and more complete physical accounts of mentality. Therefore, in general, we can level the same arguments we have used against substance dualism also against property dualism.

Evaluating the Dualist Perspective

One critique of dualism comes from the philosopher Gilbert Ryle. Ryle's (1949) argument centers on our conceptualization of mind and its relation to the body. He believes that the mind is not any particular component of the brain but, rather, all the parts

working together as a coordinated, organized whole. He illustrates with a story: Imagine a visitor from a foreign country arriving at a large university. He is shown around the campus, and the various parts of the school are pointed out to him, including the dormitories, departments, and lawns. The visitor, who has never seen any of this before, is puzzled. He says, "Well, I've seen all this, but I haven't yet seen the university." We would have to explain to him that the university is not any of the individual sites he has viewed but all the sites together and the interconnections among them (see Figure 2.3). Ryle thinks philosophers fall into the same trap as the visitor, mistaking the part or parts for the whole. He argues that the mind belongs in a conceptual category different from that of the body, just as the university is in a category different from those of the things that make it up.

Andy Clark (2001) summarizes several other critiques of dualism. These would apply to Descartes's conception as well as other perspectives. Clark says that dualism is uninformative and tells us what the mind isn't rather than what it is. If the mind isn't the brain and isn't physical, then what is it? Dualists are remarkably silent on this matter,

Figure 2.3 Where is the university?



Source: Courtesy of Peter Finger Photography (2003).

often conceding that it is something nonphysical that we can't understand yet. As a theory, dualism also is inelegant because it postulates two worlds that must be coordinated. An explanation that does not violate the principle of Occam's razor would involve a single type of world, not requiring coordination.

There are further problems with dualism. One has to do with the dependence of the mental on the physical. Factors that affect the brain, such as head trauma or drug use, have direct and dramatic mental effects. We can see that damage to a certain part of the brain—say, from a motorcycle accident—results in somewhat specific forms of mental disruption—for example, language deficits. Taking a drug such as marijuana, which alters brain chemistry, results in altered mental states. In addition, the evolutionary approach shows us that there is an approximate correlation between intelligence and the number of neurons in the cerebral cortex across species, such that species with more cortical neurons tend to have greater cognitive capacity. It is obvious from these observations that the mental is integrated with the physical, that the mind depends on the brain.

Some dualists, in response to attacks on their positions, have stated that the mind exhibits extraordinary abilities and that it would be impossible for a physical system to duplicate such abilities. For instance, how can a physical system, be it a brain or a computer, write a novel or negotiate a peace treaty? The truth is that as our technological sophistication increases, many of these abilities are becoming better understood and implemented computationally. There are now computers that can beat the best chess champions and successfully diagnose medical disorders. These are capacities once thought to be the exclusive domain of humans.

Dualists and other philosophers also argue that our subjective experiences—things such as thoughts, beliefs, and desires—are not equivalent to physical brain states. They base this conclusion primarily on introspection. When we examine what is inside our heads, they say, these subjective experiences seem to be something more than just physical. The problem with this argument is that introspection is a weak form of evidence and can be wrong (as are many of our ideas). What is required is objective proof that such experiential states are not physical.

FUNCTIONALISM: ARE MINDS LIMITED TO BRAINS?

The most influential philosophical theory of mind in cognitive science is functionalism. For this reason, we will discuss it in considerably more detail than any of the theories already discussed. To get an idea of what functionalism is about, we need to make a distinction between two ways of classifying things. **Physical kinds** are identified by their material composition only. In this view, jellyfish and carpets are different because they are made up of fundamentally different physical substances. **Functional kinds**, however, are distinguished by their actions or tendencies. Here, we could say that all automobiles fall under the same functional category because they do the same things—namely, transport goods and people—even though they may be made up of different physical substances.

So far so good, but things get interesting when we extend these ways of classifying to the idea of mind. If we think of mind as a physical kind, then minds must be the same

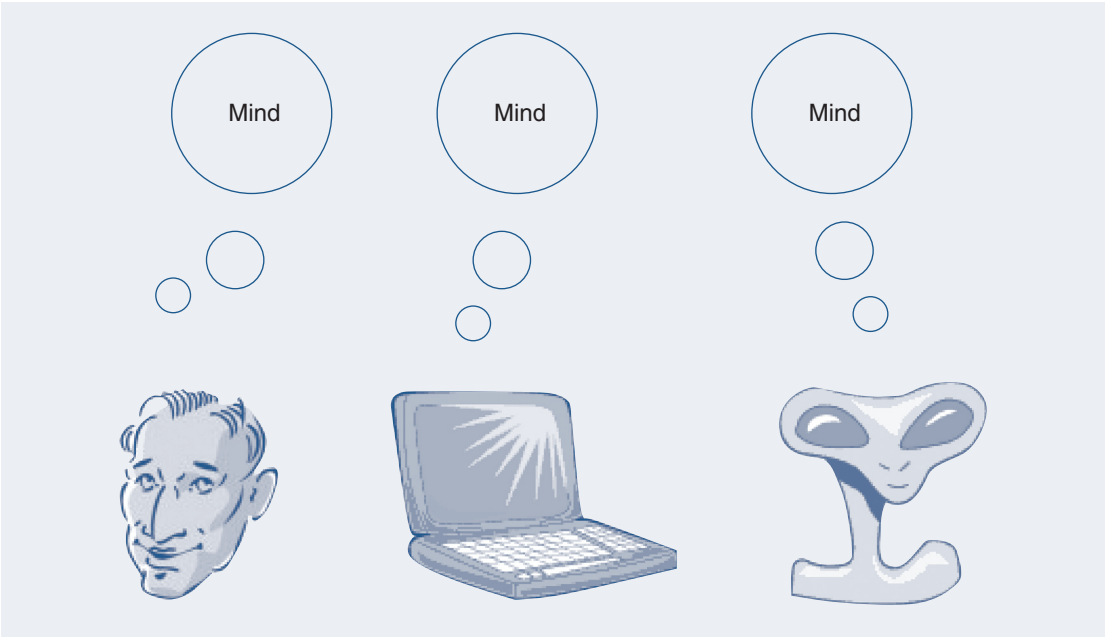
things as brains, since, as far as we know, minds cannot exist apart from physical brains. To many, this seems too exclusive. It is possible, they argue, that computers might develop minds and that there might be alien species with minds (see Figure 2.4). Neither computers nor aliens need have brains in the sense that we know them. It is more fruitful, they say, to identify minds as functional kinds and, thus, to define them by the sorts of processes they carry out rather than the stuff they're made of. According to **functionalism**, mental states are not just physical states but also the functioning or operation of those physical states. According to this view, a mind could conceivably be implemented in any physical system, artificial or natural, capable of supporting the appropriate computation.

Functionalism has several significant implications (Garfield, 1995). One is that the same mental state could be realized in quite different ways in two separate physical systems. This can be illustrated with computing devices. Two such different devices, say a desktop computer and a smartphone, can both compute the same result, such as displaying a page of text, but in different ways. The same might also be true for human computation. If we examined the brains of two people thinking exactly the same thought, we would in all likelihood not find exactly the same neural processes at work. This is the idea of multiple realization that we first introduced when discussing identity theory.

There are several schools of thought in functionalism, ranging from conservative views that advocate direct connections between physical and computational states to more liberal ones that emphasize computation over physicality. The liberal schools give two reasons for their stance. They say that for both computers and thinking organisms, the number of possible computational states always exceeds the number of possible physical states. Take, for example, all the different possible beliefs one could hold concerning politics, the environment, one's friends, and so on. Mathematically, the number of such beliefs could be practically infinite (Garfield, 1995). The number of possible physical states the brain can assume, though, is probably finite. According to that argument, the computational level of description thus becomes a richer and more diverse way of describing the mind and should be the preferred level. Second, liberal functionalists argue that psychological states such as beliefs are defined more by their relations to other such states, to inputs from the environment, and to behaviors than by their relations to physical states. A belief such as "patriotism" usually manifests itself in behaviors—for example, in flag waving. It will elicit predictable reactions to environmental stimuli—for example, feeling upset when one's country has been compromised—and will produce external behaviors such as marching or protesting. In functionalism, rather than just a brain state being what explains a patriotic state of mind, the full set of relations between the relevant thoughts and actions is what explains a patriotic state of mind.

To summarize, functionalism implies that mental states might not be reduced to any particular physical state. This argument does not quite require one to be a dualist. It is not saying that mental states don't conform to physical ones, only that there may be a wide variety of possible physical states capable of producing any given mental state. One last point of clarification needs to be made. As was the case with behaviorism, functionalism takes on two slightly different sorts of meaning within the philosophy of mind and psychology communities. Whereas the philosophical conception stresses multiple realizability, the psychological conception stresses the functions themselves. We elaborate more on this latter definition in the next chapter.

Figure 2.4 According to functionalism, different physical substrates can in principle all give rise to mind.



Evaluating the Functionalist Perspective

Although functionalism has been the dominant view in cognitive science since the 1970s, it is not without its deficiencies (Maloney, 1999). Remember that a tenet of functionalism is that minds that are not based on brains can exist. They can exist in objects such as computers as long as the physical substrates of those objects allow for the relevant computations. Critics have argued that, although it is possible that minds can exist in the absence of brains, this does not make it plausible. There is no current empirical evidence to justify this claim. We have yet to see something mental in the absence of a brain. Also, some have argued that the failure to identify mind with a physical kind can itself be considered a reason to do away with the concept of mind—rather than give it special status as a functional kind.

An additional problem with functionalism is that it cannot account for the felt or experienced character of mental states—a phenomenon known as **qualia** (quale, singular). Examples of qualia include the subjective experience of what it is like to feel hungry, to be angry, or to see the color red. It would seem that these kinds of experiences cannot be replicated as purely functional processes. A machine could be programmed to detect the color red, even mimicking the same human functional process, but this machine might not have the same conscious experience of what it is like to see red that a person has.

What is more, two individuals having the same conscious experience often do not experience it subjectively in the same way. A number of experiments have shown this to be the case with color perception. Participants looking at the same color will describe it differently (Chapanis, 1965). If asked to point out on a color spectrum what pure green looks like, one person may select a slightly yellowish green, another a slightly bluish green. This is the case even though the functional operations of their respective brains as they view the color are approximately equivalent. In this case, the neurophysiological operations behind color perception tend to be the same across most individuals.

THE KNOWLEDGE ACQUISITION PROBLEM: HOW DO WE ACQUIRE KNOWLEDGE?

A fundamental question asked by even the earliest of philosophers was, How do we acquire knowledge? Clearly, you are not born knowing everything—otherwise, you would not need to go to school and wouldn't be reading this book. But are we born knowing anything at all? Is the mind completely blank, or do we start with some rudimentary understanding of the world? One way to frame these questions is within the **nature–nurture debate**. This debate centers on the relative contributions of biology and experience in determining any particular capacity. The term *nature*, in this context, refers to traits that are genetically or biologically determined. These are coded for in our genes and so are "hardwired," meaning they are present at birth or appear at a prespecified time during development. The term *nurture* refers to traits that are learned through experience and interaction with the environment. We will examine theories of knowledge acquisition that argue for the greater influence of one or the other.

According to **nativism**, a significant body of knowledge is innate or "built into" an organism. In this sense, nativism is a theory of knowledge that favors nature over nurture. Plato was the first to outline a nativist theory of knowledge. He thought learning was a matter of recollecting what is already known—these concepts existing in the ideal world of forms and being part of our immortal souls. **Rationalism** must be subtly distinguished from nativism. Descartes was the progenitor of this perspective. Rationalists also tend to believe in the existence of innate ideas. These basic concepts include ideas such as "God" and "triangle." However, they additionally emphasize the existence of innate reasoning powers. These include certain logical propositions, such as knowing that something cannot both exist and *not* exist at the same time. We can use these a priori rational powers to form new ideas that are not given to us innately. Descartes would agree that we are not born with the idea of "table" but can acquire it given our innate ability to perceive, think about, and interact with objects. By applying formal logic to the propositions we assume are true, rationalism allows us to combine them to generate new knowledge that must also be true.

For example, let us assume the following premises are true:

1. If a person cheats on their taxes, that person is a criminal.
2. Dan cheats on his taxes.

Aristotelian logic allows us to generate a new conclusion for this syllogism and add to our list of premises-assumed-to-be-true the following conclusion:

3. Dan is a criminal.

Implied in this particular logical syllogism is a kind of causal relationship between tax cheating and being a criminal. But sometimes the directionality of the causal relationship can get confusing. Occasionally, human reasoners will simply treat Premise 1 as though it is a bidirectional correlation, such that not only are tax cheaters criminals, but criminals are also tax cheaters. In such instances, if Premise 2 were replaced with "Dan is a criminal," then a person using that flawed interpretation of Premise 1 might conclude that "Dan cheats on his taxes." This is called making the logical error of *confirming the antecedent*. Just because Dan is a criminal doesn't allow us to use Premise 1 to conclude that he is a tax cheater. While all tax cheaters are criminals (because cheating on one's taxes is a crime), some people fail to consider the counterfactual (alternative) possibility that some criminals may not cheat on their taxes—they just commit other crimes. Therefore, not all criminals are tax cheaters. The directionality of the causal relationship might be easier to identify in this next example.

Let us assume that Premises A and B are true:

A. If one smokes cigarettes, one increases one's likelihood of getting cancer.

B. Joe smokes cigarettes.

Like before, we can apply logic to generate a new conclusion:

C. Joe is increasing his likelihood of getting cancer.

The directionality of the causal relationship in this example is easier to see because we know that multiple things can cause cancer (not just cigarettes). So if Premise B were replaced with "Joe has cancer," most people would not falsely confirm the antecedent and jump to the conclusion that Joe is a smoker. (Although they might ask.) When reasoning about causality, it is important to both consider counterfactual circumstances (such as "If Joe is not a smoker, could he still get cancer?") and to avoid confirming the antecedent.

In contrast to nativism (and rationalism), **empiricism** sees knowledge as acquired through experience and observation: It favors nurture over nature. Rather than relying on knowledge that you might start out with (e.g., a list of premises-assumed-to-be-true) and then using rationalism to logically combine them and generate new knowledge, empiricism requires new knowledge to be acquired by testing your hypotheses in the world. For instance, if you think Joe in the example above might have cancer, then run some medical tests on him. Or if you think Dan might be a criminal, then test that

theory by collecting evidence and presenting it in a court of law. In this view, knowledge gets into the head through interaction with an environment, meaning it is learned. The senses provide the primary channels via which knowledge of the world is generated. Our knowledge of the concept "lemon" in this account begins with looking at a lemon, touching it, and tasting it. The British philosopher John Locke (1632–1704) is credited as the founder of the empiricist movement. He used the phrase *tabula rasa*, which literally translates as "blank slate" (essentially a chalkboard with nothing written on it). Locke believed that we are born as blank slates, lacking any knowledge, and that over time experience puts writing on the slate, filling it up.

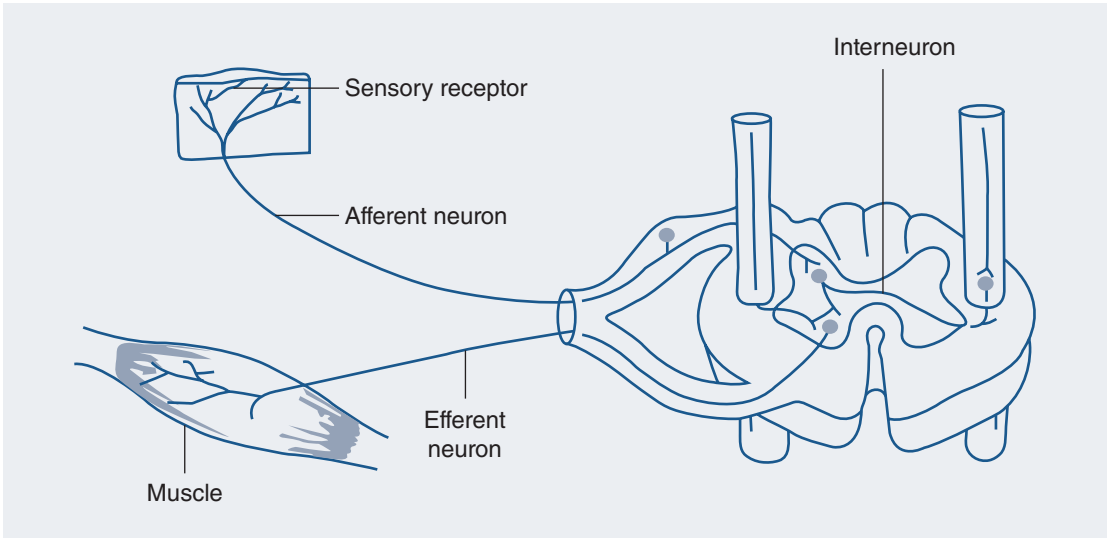
Locke had a more fully developed theory of learning. He differentiated between simple ideas and complex ideas. **Simple ideas** are derived through sensory input or simple processes of reflection. They are received passively by the mind and cannot be reduced to simpler ideas. Looking at a cherry would generate the simple idea of "red." Tasting a cherry would produce the simple idea of "sweet." **Complex ideas** are formed from the active mental combination of simple ideas. They are created through reflection only and can be reduced to parts, their component simple ideas. The idea of "cherry" would result from the associative combination of such simple ideas as "red," "sweet," and other commonly occurring sensations derived from one's experiencing cherries. This cluster of simple ideas is naturally associated because each time we experience a cherry, many of those simple ideas are generated. For this reason, Locke and others who have proposed similar notions are sometimes known as the associationists.

Evaluating the Knowledge Acquisition Debate

One might be tempted to immediately dismiss the doctrine of innate ideas put forth by the nativists and rationalists. After all, it seems absurd that we should be born knowing factual information such as the content of the Gettysburg Address. But the scope of knowledge is broader than this. Think back to the previous chapter, in which we defined declarative knowledge for facts and procedural knowledge for skills. There is quite a bit of research supporting the notion that some forms of procedural knowledge are innate. Newborn infants, for instance, come into this world with a variety of different skills. These skills are universal across the human species and manifest themselves so soon after birth that they couldn't possibly have been learned. Therefore, they qualify as examples of innate knowledge. Let us examine a few of these.

All healthy infants demonstrate a set of **reflexes**. These reflexes include the grasping reflex, in which the fingers tighten around a touch to the palm, and the rooting reflex, in which the infant turns his or her head and begins sucking an object placed near the mouth. Reflexes serve a clear adaptive function. Grasping and sucking, along with behaviors generated by other early reflexes, are important for survival. The physiology behind reflexes is simple and fairly well understood. A stimulus triggers one or more sensory neurons that then activate intermediary neurons. These in turn activate motor neurons, causing the resulting behavior. It is easy to see how such a simple mechanism could be hardwired at birth to enable the infant to respond effectively to its environment. Figure 2.5 shows the anatomy of a spinal reflex.

Figure 2.5 The neural connections in a spinal reflex. A reflex is an example of innate procedural knowledge.



Smell preference is another example of innate behavior. Steiner (1979) found that newborns tend to agree with adults in terms of which odors they consider pleasant or unpleasant. He found that odors such as strawberry and banana elicited agreeable facial expressions from young infants—for example, smiling. Unpleasant odors, such as rotten eggs, elicited expressions of disgust. As is the case with reflexes, these smell preferences have survival value. Babies who find the smell of fruit attractive will eat the fruit and thereby gain nutrients; those who are repulsed by spoiled food will reject the food and avoid getting sick. The neural mechanisms behind such preferences are probably not too complex either. They need to involve little more than a mapping between the odor and the emotional response.

Given the above examples, we see that it is not so far-fetched for us to be born with some basic procedural knowledge. This knowledge is in the form of simple neural circuits that map stimulus inputs to appropriate behavioral outputs. This knowledge can even be represented using the conditional rules we talked about in Chapter 1. A computational coding of a smell preference might look something like this: "If *smell is rotten eggs*, then *disgust*." The odor, if it satisfies the first part of the conditional, would then trigger the response in the second part.

But how did these circuits get there in the first place? The early nativists and rationalists either did not specify the source of innate knowledge or attributed it to God. Evolutionary psychology offers us another explanation. It attributes such capacities to generations of selection pressures acting on a species. Basically, over many generations,

individuals who happened to not be born with those disgust-based procedural rules found themselves eating things that led to them dying before they could contribute their DNA to the next generation. Those who were born with those rules succeeded in contributing their DNA to the next generation. The result is that eventually the entire species is generally born with those procedural rules innately programmed into their nervous systems. These pressures can promote the development of adaptive (survival-related) cognitive abilities. Evolutionary psychologists can be considered modern-day nativists. See the evolutionary approach chapter for more on their views.

The phrasing of the nature–nurture debate as allowing for only a single alternative, either one or the other, is misleading. Although some traits may indeed be mostly the product of nature or of nurture, there is a large middle ground consisting of traits or cognitive abilities that can result from the complex interaction of the two. In these cases, nature may set constraints or limits on environmental influence. Take memory, for example. Tsien, Huerta, and Tonegawa (1996) engineered a mutation in a gene that affects a particular type of receptor in the hippocampus, a brain area responsible for the learning of new information. Rats with the mutation did poorly in a memory task as compared with normal rats in a control group. Tang et al. (1999) did something even more remarkable. Through genetic manipulation, they increased production of a particular subunit in the hippocampal receptor. This change increased the receptor's effectiveness. Rats with this "enhanced" version of the gene outperformed rats with normal receptors on a spatial memory test.

This research is exciting because it shows that memory in these animals is at least partially under genetic control. However, it is also well documented that human memory capability can be improved through training and the use of memory strategies (Roediger, 1980). The way in which these genetic and environmental factors interact to determine memory in any given individual is sure to be complex. A person with an innately poor memory could probably improve their memory skills to better than average by using a variety of memory strategies. And a person with an innately strong memory could make their memory skills look almost superhuman by using those memory strategies.

THE MYSTERY OF CONSCIOUSNESS: WHAT IS CONSCIOUSNESS AND HOW DOES IT OPERATE?

Consciousness is a complex concept and has no single agreed-on definition. In its broadest sense, we can think of it as the subjective quality of experience (Chalmers, 1996). It may be thought of as our individual subjective awareness of mental states. These states include sensation, perception, visual images, conscious thought processes, emotions, and sense of self, just to name a few. But these states assume that a person is in a normal, awake, and alert frame of mind. The issue becomes more complex when we think of other types of consciousness—for example, being unconscious, asleep, in a drug-induced state, hypnotized, or meditating. There are clinical cases representing other states of consciousness as well. In dissociative identity disorder, a person can alternate between separate personalities. Each personality can possess unique skills and may or may not be aware of the others. In split-brain patients, one half of the brain can

possess an awareness of an object that the other half does not possess. For simplicity, we do not consider these alternate states of mind.

An interesting aspect of consciousness is whether it is unitary or divided. Subjectively, our consciousness seems to be unitary. That is, one recognizes himself or herself to be one person, experiencing things in the present moment. When one studies the brain, though, one finds that there is no single place or even time where consciousness seems to happen. Instead, the brain in action is a case of activity going on all over the place. Furthermore, the brain may even be processing different aspects of a single experience at different times. How can we reconcile this objective evidence with our subjective experience? See the "Interdisciplinary Crossroads" section for one theory on this apparent contradiction.

Chalmers (1996) makes a distinction between phenomenal and psychological concepts of mind. The **phenomenal concept of mind** is essentially the idea of mind as a conscious experience. Mental states in this view need to be explained in terms of how they feel. The **psychological concept of mind** sees mental states only in terms of how they cause and explain behavior. Here, mind is characterized by what it does—how it feels is irrelevant. Philosophers have concerned themselves primarily with the former, psychologists and cognitive scientists with the latter. To make this distinction clear, imagine biting into a candy bar. A phenomenal investigation would attempt to explain why you experience the mental states of "sweetness" or "chocolate" and why you might perceive them differently than somebody else does. A psychological investigation would concern itself with the neural circuits that become activated during the taste, how they might be represented computationally, and how this explains when you might stop eating. In this section, we concern ourselves with the phenomenal concept of mind and its relation to consciousness, since the psychological view is in most cases the topic of the remainder of this book.

Chalmers (1996) also differentiates between what he calls the easy and hard problems of consciousness. **Easy problems of consciousness** are those that can be solved by cognitive science and explained in terms of computational or neural mechanisms. His examples include the ability to discriminate, categorize, and react to environmental stimuli; the focus of attention; and the difference between wakefulness and sleep. Obviously, these correspond to the psychological concept of mind. The **hard problem of consciousness** involves subjective experience. Here we would need to explain why we have a visual or auditory experience when we look at or listen to something. In this case, we are now talking about the phenomenal concept of mind. Whereas science can give us answers to the easy problems, it may not be possible to provide answers to the hard problem of consciousness. The fact that subjective human experience may not be fully explained by an objective account using physical and mechanical processes is known as the **explanatory gap** (Levine, 1983).

The What-It's-Like Argument

Nagel (1974) says that there is "something that it is like" to have a conscious mental state. When you bite into a candy bar, you have a subjective conscious experience of

tasting it. The candy bar, of course, has no such experience. There is nothing that "it is like" for the candy bar being bitten. This is one way of describing consciousness—that organisms that possess it can be described as having some sort of experience. Things incapable of supporting consciousness cannot.

But what is this experience like? Nagel (1974) asks us to imagine what it must be like for a bat to navigate by echolocation. In echolocation, the bat emits high-pitched sounds. The sound waves bounce off an object in the animal's path, and the animal uses the reflection time as a measure of the object's distance (see Figure 2.6). Neuroscientists have developed an in-depth understanding of the neuropsychological mechanisms that allow a bat to perceive its environment this way (Ulanovsky & Moss, 2008). And engineers have built adaptive sonar machines that can compute echolocation in a way similar to how a bat does. But none of this tells us what it is like for the bat to experience the world in the way it does. We have seen this argument before in our evaluation of functionalism. There, we said that a functional description of a cognitive process does not account for the qualia, or subjective experience, of the process.

Figure 2.6 What is it like to be a bat?



The problem here is that science can provide only an objective account of a phenomenon, and consciousness is an inherently subjective state. As organisms capable of supporting consciousness, we can introspect and analyze what it is like to have or to experience a mental state. Unfortunately, for cognitive science, this is not what is needed. Cognitive science must, instead, have a scientific and objective account of what consciousness is. Frank Jackson (1982) aptly illustrates the difference between objective and subjective accounts of a conscious experience. He asks us to think about a neuroscientist named Mary who is well trained in the physical mechanisms underlying color vision. She understands everything there is to know about how the eye and brain process color information. Mary, however, is colorblind. Imagine now that we take away her colorblindness and allow her to look at a tomato. Interesting questions arise. Does Mary learn anything new by this experience? Does the scientific community gain anything by Mary's (or anybody else's) description of what it is like to see red? Jackson argues that we do gain something and that science needs to explain this new information.

In contrast to this position, some state that subjective knowledge is not factual knowledge at all and, therefore, does not constitute any kind of an explanation. Personally knowing what it is like to taste a candy bar or to see red is not the same thing as objectively and factually knowing it. Adopting this position, we as investigators would be forced to ignore introspection and any other form of subjective description. Our focus would be on only legitimate objective techniques for studying the mind, such as observation, experimentation, brain imaging, and computer simulations.

Mind as an Emergent Property

Consciousness is a "hot" topic in contemporary cognitive science. In the past 15 or so years, there have been renewed interdisciplinary efforts to understand it. A number of well-known authors have published books for academic and layperson audiences that outline their definitions and views on the subject. These authors' theories are too numerous to mention here. We instead describe one popular theory in this section and a more detailed description of another in the "Interdisciplinary Crossroads" part of the chapter. Before we begin this, however, we must articulate two different conceptions of how complex systems such as the brain might work.

Reductionism is the belief that everything about a whole can be explained entirely by understanding its parts. If this were true, we could say that Phenomenon X is "nothing more than" Phenomena Y and Z combined. If it was a simple additive combination of $X = Y + Z$, then the reduction would be easy, and talk of X could be eliminated because $Y + Z$ tells us all we need to know. If we were to let X be understanding the mind, Y be the connectivity among neurons, and Z be the activation of those neurons, then $X = Y + Z$ would mean that we could completely understand the mind by a linear combination of neural connectivity and neural activation. Would that it were so easy. Few, if any, natural phenomena actually lend themselves to a linear additive reduction like $Y + Z$. Instead, the subcomponents that combine to generate some coherent event, like Phenomenon X, are usually combined in rather nonlinear (nonadditive) ways that involve processes akin to multiplication, division, and exponents. There are very few cases of successful linear reduction in the sciences. Thermodynamics has not been fully

reduced to statistical mechanics, nor has chemistry been fully reduced to quantum mechanics. In part, this is because it is not enough to understand the parts; one must also understand how they interact or relate to one another.

In **emergence**, the features of a whole are not completely dependent on an additive combination of the parts that make it up. Some of those features emerge over and above what the additive combination would predict. Holland (1998) lists several characteristics of emergence. Emergent systems are made up of interacting parts. The function of these systems is rule governed. The rules stay the same even though the parts change over time. And the emergent system's output often contributes to its next input. This feedback loop gives rise to novelty and unpredictability. It is difficult or impossible to anticipate what such systems will do, even if some or all of the rules that govern their operation are known. A problem with the emergent account is that it doesn't constitute an explanation. All it says is that the behavior of the whole is more than the sum of its parts. What is needed is a causal or scientific explanation of how part interactions give rise to emergent features.

John Searle (1992) uses the concept of emergence in his book *The Rediscovery of the Mind.* He argues that consciousness is an **emergent property** of the brain. An emergent property of a system, as we have mentioned, is realized through the interaction of the system's parts. Searle says if we have a given emergent system, S, made up of elements a, b, c, and so on, then the features of S may not be the same as the features of a, b, c, and so on. This is because the features of S arise from the causal interactions of the parts. Water, for example, has the features or properties of liquidity and transparency. The H_2O molecules that make it up do not share these properties. The causal interactions of these molecules give rise to these properties. In the same way, Searle says, consciousness is a property of the brain but not of its parts. If we take neurons to be the relative parts, then they have their own properties, such as being able to communicate with one another via electrochemical signals. These properties that are inherent in the way the neurons interact give rise to consciousness, but the properties of individual neurons need not be those of a conscious mind.

Searle is very careful to point out that he is not a reductionist. He does not believe consciousness is reducible to its parts. In reductionism, explanation goes downward and a phenomenon is directly explainable in terms of what is happening at a smaller scale. In emergence, explanation goes upward. The large-scale phenomena are more than just what is happening in and around the small-scale parts and cannot be explained solely by an account of what the parts are doing. This idea is similar to the concept of a gestalt in perception. Gestalts are discussed in the next chapter ("The Psychological Approach").

Searle seeks to avoid the monism–dualism dichotomy of the mind–body problem. He does this by talking about consciousness as a property rather than a substance. He likens consciousness to an emergent characteristic of what brains do in the same way that digestion is what stomachs do or photosynthesis is what plants do. He sees consciousness as a natural process and a by-product of the brain's nature. However, he does classify conscious mental states as separate from physical ones. He states that they constitute a unique and novel category of phenomena, with an independent reality and a distinct metaphysical status.

EVALUATING THE EMERGENT VIEW OF MIND

As appealing as this formulation is, it still leaves us with some vexing questions. The reformulation of consciousness as a property, and a nonphysical one at that, still begs the question: What is a property? If a property is not physical, then of what substance is it? Although attempting to avoid the mind–body debate, Searle seemingly ends up as a property dualist. Restating consciousness as a nonmaterial property of a material brain doesn't get us any further toward understanding what this type of property is. Also, it is not clear how emergence happens—that is, we do not yet have an understanding of the relationship between microscopic and macroscopic properties. In the case of water, we can say its large-scale properties have something to do with the three-dimensional shape of the H_2O molecules and other conditions, such as the surrounding temperature. For consciousness and the brain, this relationship between the microscopic and the macroscopic is far more ambiguous.

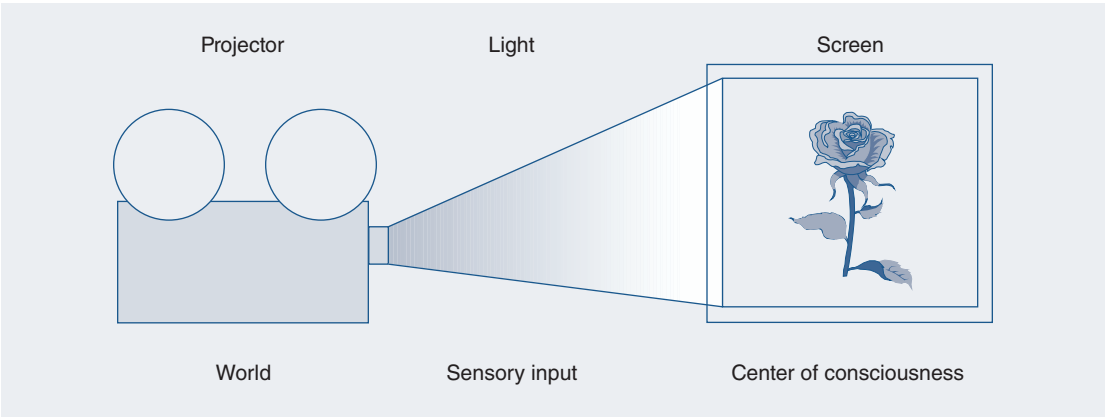
Searle's reason for believing in a nonmaterial consciousness is based on his conception of the difference between physical and mental things. For physical things, we can make a distinction between appearance and reality. A piece of wood may subjectively appear a certain way to us—as brown, as having a certain length and weight, and so on. These characteristics can also be measured objectively; we can put the wood on a scale to determine its weight, use a ruler to determine its length, and apply a wavelength detector to measure its color. For mental things, this distinction between the subjective and the objective goes away. Regarding mental experience, Searle believes that appearance is the same as reality and that our subjective introspections are objectively correct. But if this were true, we would have to trust our intuitions about the mental world as metaphysically special and nonmaterial.

In opposition, Paul Churchland (1995) points out that this reliance on the infallibility of introspection is an outdated notion. He notes that introspection often does not give us direct and accurate knowledge of the mental. Our inner assessments of mental states can be quite often and, notoriously, wrong. It is commonplace for us to err in judging our thoughts, feelings, and desires. Many of the early psychologists relied on introspection as a means to study the mind. The next chapter provides a more elaborate discussion of the problems they encountered.

Consciousness: One or Many?

In his book *Consciousness Explained*, Dennett (1991) outlines an interesting theory on the nature of consciousness. He begins by refuting the classical view of consciousness. The classical view, promoted by Descartes, posits a single place in the brain where all information funnels in. This area is a supposed center of consciousness, where we experience the world or the contents of our thoughts in a coherent, unified way. Dennett calls this center the "Cartesian theater." It is as though our consciousness is the result of a projector displaying information on a movie screen. The individual sitting in the theater watching the screen then has a single conscious experience of what is playing. Figure 2.7 gives a representation of the Cartesian theater.

Figure 2.7 The Cartesian theater explanation of consciousness.



There are a number of problems with the Cartesian theater. To start, linked modes of information do not arrive within the brain simultaneously. Light from an event precedes the arrival of sound. The sight of a fireworks display reaches the mind prior to the sound of the explosion, yet we experience the two in unison. This suggests that our consciousness is constructed; the visual experience is kept in check or delayed until arrival of the sound, at which point the two are integrated into a unified percept of the fireworks. This example and others imply that consciousness does not occur in real time—but (in many instances) several fractions of a second or so after an event. Our experience of consciousness as direct and immediate seems to be an illusion.

Another problem with the Cartesian theater is that anatomically, it is difficult to find a solitary brain region that, all by itself, links incoming sensory inputs and outgoing motor outputs. There is no **central processing unit (CPU)** in the brain as there is in a computer. The task of a computer's CPU is to schedule and coordinate ongoing activity. Furthermore, the Cartesian theater analogy requires an observer in the audience watching the screen. This observer is the subjective self who experiences the screen's contents. But how is this person inside one's head interpreting the image and having the conscious experience? To explain this, we would need to posit another mechanism or theater inside this person's head with another even smaller person and so on, ad infinitum. This is known as the homunculus problem in psychology and philosophy. **Homunculus** translated means "little man." An effective theory of consciousness must avoid the logical conundrum of homunculi nested inside each other.

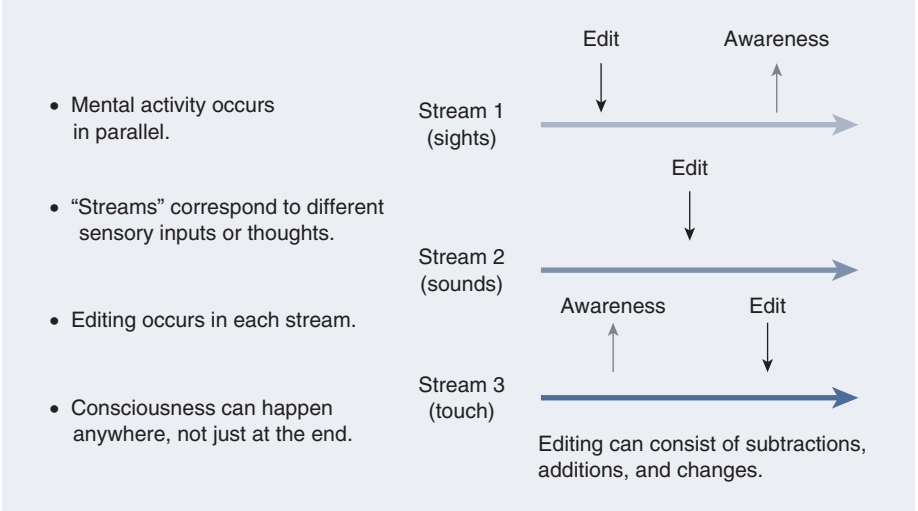
Dennett replaces this problematic formulation with a multiple-drafts model of consciousness (see Figure 2.8). In this model, mental activity occurs in parallel. Rather than projecting to a single location for processing in unison, different ongoing streams of information are processed at different times. Each of these streams can correspond

to different sensory inputs or thoughts. Processing or editing of the streams can occur, which may change their content. Editing can consist of subtractions, additions, and changes to the information. Awareness of a stream's content can happen before or after editing takes place. To illustrate, take our fireworks example. One mental stream would contain the visual experience of the fireworks, while another would contain its auditory representation. The visual stream would undergo editing in the form of a delay to synchronize it with the auditory stream. Then the information from both streams could be tapped to produce awareness.

There is abundant evidence in support of the multiple-drafts model. Take, for instance, the organization of the visual system. It adopts a "divide-and-conquer" strategy. The visual system carves up different aspects of an object during pattern recognition. These aspects are each processed separately in different parts of the brain by anatomically distinct pathways. This information is later combined to yield a unitary percept, but we are not aware that the information has been separated and then united.

A famous experiment by Loftus and Palmer (1974) also provides support for Dennett's idea. In this study, participants viewed a film of a car crash. Afterward, they were asked to estimate the speeds of the cars. The crucial manipulation was in how the question was asked. Some were asked how fast the cars were going when they "bumped" into each other. Others were asked how fast they were going when they "smashed" into each other. As you might predict, those who were queried with a mild descriptor such as "bumped" estimated that the cars were moving more slowly. Those queried with a more severe descriptor such as "smashed" estimated the speeds as considerably higher. In fact, even simply asking "what was happening?" in the event—instead of asking "what happened?"—can elicit more violent descriptions of the car crash. Essentially, the past

Figure 2.8 Dennett's multiple-drafts model of consciousness.



progressive verb form places your memory's perspective in the middle of the event, whereas the past perfect verb form places your memory's perspective outside of the event (Matlock, Sparks, Matthews, Hunter, & Huette, 2012). These results suggest that our memories of an event are not perfectly preserved "snapshots" of what happened but are actively edited over time. Subtle differences in the posing of a question, and other subsequent experiences after the event, can cause the memory to be edited and changed.

Dennett's theory also allows for different levels of awareness. Some information that is part of a stream may be available to conscious awareness and could be verbally described by the individual experiencing it. Other data streams we may be only vaguely aware of, but they can persist and influence additional mental processes. Yet other information may simply fade into the background. We may never be aware of this information. These three levels of awareness are comparable to Freud's conscious, preconscious, and subconscious aspects of mind, discussed in the next chapter.

In summary, Dennett's theory is more logically coherent and captures some of the empirical evidence on conscious experience. It suggests that there is no central place where consciousness happens but that multiple mental events occur in parallel. These events may be edited and changed in such a way that consciousness need not take place in real time. We may or may not be aware of these events.

Consciousness and Neuroscience

What does the brain have to do with consciousness? Is there some part of the brain or some particular pattern of neural activity that gives rise to consciousness? What is the neural correlate of conscious experience? Although philosophers have been debating the relation between the brain and mental phenomena for millennia, recent advances in neuroscience have yielded more specific insights into these questions. Let's examine some of them here.

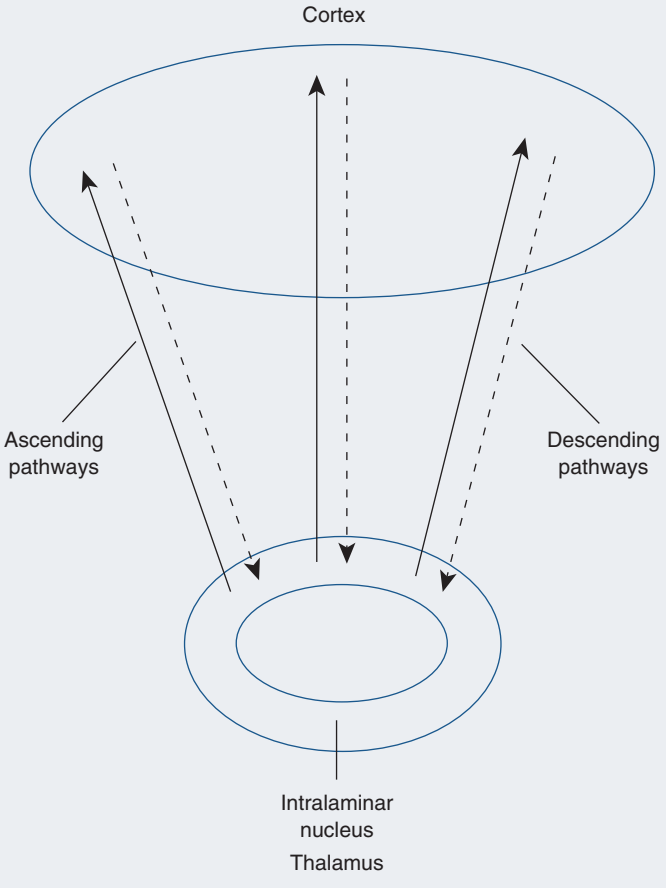
In general, the neuroscience view is that consciousness results from the coordinated activity of a population of neurons in the brain. Popper and Eccles (1981) see consciousness as an emergent property of a large number of interacting neurons. A different idea is that there are neurons specifically devoted to producing consciousness. Crick and Koch (1995) believe that these are located throughout the cortex and in other areas associated with the cortex. Activity in at least some subset of these neurons produces conscious experience. They believe that these neurons are special and that they differ from other neurons in terms of their structure and function. A similar but slightly different conception is that any cortical neuron may contribute to a conscious experience; however, different groups of cortical neurons mediate different types of conscious experience.

If there were special consciousness neurons, where might they be located? It has been proposed that one area is the intralaminar nuclei of the thalamus (Purpura & Schiff, 1997). The thalamus is a relay center for incoming sensory information. It sends information from each of the different sensory modalities, such as vision, audition, touch, and taste, to specialized areas of the cortex devoted to processing the information. Lesions of or damage to this brain region results in coma and loss of consciousness. It may be that these thalamic neurons, because they have widespread projections to many cortical areas, serve to activate or arouse other cortical neurons. Specific activity in different cortical

regions may then account for specific forms of consciousness. For example, activation or arousal in the occipital visual regions may correspond to visual awareness, while activation of the somatosensory cortex may produce awareness of different parts of the body.

Churchland (1995) formulates a neurocomputational theory of consciousness that focuses on connections between the intralaminar nuclei of the thalamus and disparate cortical areas. The circuit consists of ascending projections from the thalamus to the cortex, as well as descending pathways from the cortex to the thalamus. Figure 2.9 shows the anatomical layout of this area. These pathways are recurrent, meaning that a signal can be sent back and forth inside it. In this case, information coming into the thalamus

Figure 2.9 The consciousness circuit proposed by Paul Churchland. Activity in these pathways may give rise to conscious experience.



can be passed to the cortex, while the cortex can also pass information back to the thalamus. Recurrence is an important network property because it allows for feedback and learning. Recurrent activity in a network may sustain information over time and be the basis for conscious mental awareness. Recurrent artificial neural networks and their properties are described in Chapter 7 ("The Network Approach").

Churchland believes that the characteristics of this network can account for a number of different features of consciousness. One such feature is the capacity of consciousness to hold information over time—the equivalent of a short-term memory in which we are aware of the current moment in relation to the past. This is in keeping with the network's recurrent nature, since information can be held over time as it is cycled back and forth. Churchland also shows that this network can maintain activity in the absence of sensory inputs—for example, when we are daydreaming or thinking with our eyes shut. It can additionally explain why we lose consciousness during sleep, why it reappears during dreaming, and a host of other such features.

Churchland is quick to acknowledge, however, that it is the dynamical properties of this recurrent network and not its particular neural locus that make consciousness possible. He admits that a consciousness circuit may exist in places that have been suggested by other researchers. One such area is within the right parietal lobe (Damasio, 1994). The neuroscientist Rodolfo Llinas (2002) has suggested that consciousness may arise within the layers of the primary sensory cortex itself. He has written a comprehensive book that lays out his views.

Koch (2004) has also adopted a neurobiological approach to consciousness. He defines the **neural correlates of consciousness (NCCs)** as the minimal set of neural events and structures sufficient for a specific conscious experience. Since the entire brain appears to be sufficient for consciousness, the term *minimal* in the above definition is used because we would like to know what specific brain areas or activity is necessary. There should be as many distinct NCCs as there are different types of conscious experience. For instance, one pattern of activity may underlie the taste of chocolate, another the feel of wind on your face, and yet another the feeling of jealousy. If we change this pattern of activity, there should be a corresponding change in the subjective experience, and if we suppress it, the associated experience should disappear. Furthermore, it should be possible to induce a particular experience artificially by creating an NCC through an electrode or some other form of stimulation.

INTERDISCIPLINARY CROSSROADS: PHILOSOPHY, NEUROSCIENCE, AND BINOCULAR RIVALRY

Imagine an experiment where you are asked to put on glasses with one red lens and one blue lens. On a screen in front of you, there is a pattern consisting of vertical red bars crossed by

(Continued)

(Continued)

horizontal blue bars. The glasses present a different image to each eye: One sees the red stripes; the other the blue stripes. Under normal circumstances, the information from both eyes is fused to form a single unitary sense of the world with depth. Under these new circumstances, though, the two eyes compete for conscious awareness. You might first see the horizontal blue bars for a few seconds; then there would be a shift, and for the next few seconds, you would perceive the vertical red bars. This phenomenon, where each perception alternately vies for visual consciousness, is called **binocular rivalry** (Alais & Blake, 2005).

Now in this study, we ask you to push two buttons, each corresponding to one of the possible percepts. You are instructed to push the left button when you see the vertical red bars and the right button for the competing perception. While this is happening, the researchers are measuring your brain activity using an imaging technique known as magnetoencephalography. In this procedure, a helmet containing a number of sensors that detect very small magnetic fields induced by neural activity is placed over your head. This allows the investigators to see which parts of your brain are active during the experience. In this way, they can determine the neural correlates of these conscious experiences.

The results from this work show a particular pattern of neural action when participants reported not being consciously aware of the images. This pattern is characterized by widespread action in the visual areas of the occipital lobes as well as in the frontal lobes that underlie more high-level cognitive processing. This pattern, however, changed dramatically once an observer reported seeing one of the two possible sets of bars. Awareness of the pattern produced an increase of 40% to 80% neural responding. There was much reentrant activity in this state, with looping feedback found between different brain regions. But perhaps what is most interesting is that in these studies, no two observers had the same pattern, even when they reported seeing the same thing. These sorts of individual differences pose a problem for cognitive science because they imply that the brain does not always rely on the same areas to produce a given function. Fortunately, brain imaging does show some localization of function across individuals, as well as variability.

In another similar study, participants were presented a picture of a house to one eye and a picture of a face to the other eye (Tong, Nakayama, Vaughn, & Kanwisher, 1998). When the observers reported seeing a house, functional magnetic resonance imaging showed bilateral activation in the parahippocampal place area (PPA) and less activity in the fusiform face area (FFA). When they perceived the face, this pattern of activity was reversed, with greater activation in the FFA and less in the PPA. The FFA, as we will describe in more detail in the neuroscience chapter, is that portion of the temporal lobe that seems to respond selectively to facial stimuli.

Experiments such as these can tell us what parts of the brain or what patterns of brain activity underlie conscious experience. Ultimately, the use of such imaging techniques may be able to give us a complete picture of brain action for any given conscious experience in any individual. This will certainly tell us a lot about the brain, but it falls short of explaining the subjective nature of conscious experience. That is because, in the end, we will be left with an objective description of an inherently subjective phenomenon. In fact, that is all that science as an objective method can provide. At that point, we will either need to be content with our physicalist account of consciousness at the neural level or continue our investigation further by looking for more physical correlates at different levels, such as the synaptic or molecular.

Consciousness and Artificial Intelligence

Researchers in AI design algorithms to perform real-world computational tasks such as language comprehension and problem solving. Many of these algorithms can be judged as successful from a behavioral standpoint because they adequately perform their tasks, some under a variety of different conditions. If we define thought as computation in some physical substrate, as functionalists do, then we can also, without much risk, say that the execution of these programs is equivalent to "thinking." But does this correspond to consciousness? This is a much riskier proposition, since, as we have seen, consciousness can imply more than computation. It seems to involve subjective experience and perhaps other things. In this section, we address the question of whether a machine can be conscious. This is perhaps the most interesting philosophical issue in AI today.

There are a variety of different perspectives on whether or not a machine can become conscious (Freedman, 1994). These may be generally classified into two categories. The **strong AI** view asserts that consciousness can arise from a purely physical nonbiological process. Followers of this perspective believe that, eventually, as we create machines with greater complexity and computational power, we will see consciousness emerge in them. Proponents of **weak AI** claim that consciousness is itself either not a physical process, and so can never be reproduced, or a physical process that is so complex that we will never be able to duplicate it artificially.

Let us examine the arguments both for and against strong AI. Daniel Dennett (1998) raises several points in its defense. He mentions that many phenomena that used to have mystical and supernatural explanations now have scientific ones. Consciousness should be no different, he argues. Some have claimed that consciousness may be possible only in an organic brain. Dennett concedes that this may be true but notes that science has already been able to mechanically reproduce small-scale biochemical processes. An alternate counterargument is that consciousness is simply too complex to be artificially replicated. In response to this, Dennett says that consciousness of a more basic form may not require a sophisticated artificial substrate. Dennett ends by noting that any conscious machine will probably have to develop this capacity through an extended learning process, just as humans do. From a practical standpoint, this is not a barrier, since a number of machines that learn from experience have been designed.

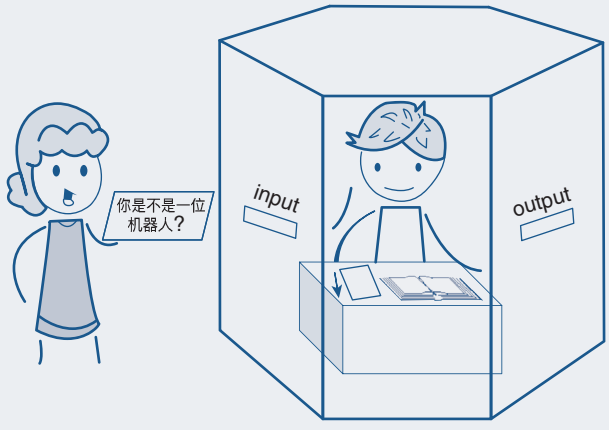
Perhaps the most persuasive and well-known argument against the strong AI position is the **Chinese room scenario** (Searle, 1980). In this hypothetical situation, a man who doesn't speak or read Chinese is in a room by himself. Outside the room is a person who writes a question or comment in Chinese and slides the paper through an input slot to the room. Although the man in the room understands no Chinese whatsoever, he has a book of rules that tells him how to relate the Chinese symbols on the incoming paper into a set of symbols constituting a suitable response (see Figure 2.10). He then traces those symbols on another piece of paper and sends it through the output slot. For example, if the outside person utters, "How are you?" in Chinese, the man in the room may, using the rule book, counter with, "I'm doing fine" in Chinese. To an outside observer, it would seem as if the person in the room understands Chinese. After all, he has a fluent reply to any question or comment that is given. But Searle's point is that the man knows

no Chinese and therefore has no understanding of the content of the conversation. He is only following a prescribed set of rules that maps one set of symbols onto another. Those symbols that he is using are not grounded in the world for him; they have no meaning for him. This is a rote execution of an algorithm, and according to Searle, it is all a machine can do. Therefore, he says, machines can never "understand," "know," or "be aware of" the information they process. They cannot be conscious of what they do. Consciousness of the human sort requires something more than just following an algorithm. To Searle, these extra ingredients are intentionality and meaning—aspects of mental representation discussed in the introductory chapter.

Boden (1990) raises a number of objections to the Chinese room argument. First, the terms *understanding* and *intentionality* are not well defined. Understanding could be operationally defined as being able to respond successfully when asked, rather than as entailing some inherent meaning on the part of the person. Second, a person who remained for some time in the Chinese room (or even a machine in the room with sufficient complexity) might eventually achieve some level of understanding. Either might eventually notice that certain combinations of characters always go together and from this learn the basic elements of syntax.

Moreover, Andy Clark's (2008) proposal that the informational tools you use become *part of your mind* (not just appendages for your cognition) would suggest that while any

Figure 2.10 Searle's Chinese room scenario.



one part of the Chinese room may not look like it understands Chinese, the room as a whole system does process and understand Chinese. Consider when you use your smartphone to have a text-based conversation with friends. Your brain knows English pretty well, but it may not always have the correct spelling for certain words, it probably hasn't memorized every single one of the emojis at your disposal, and it certainly cannot send a text message by itself. Your thumbs have learned to move in quick, articulated ways that work well with the smartphone interface, but they too cannot send a text message by themselves. And your smartphone can make loads of clever suggestions for upcoming words or emojis, and it autocorrects your misspelled words, but it better not be sending a text message by itself! Your brain, your thumbs, and your smartphone together form a whole system that fully understands how to process a text-based conversation overloaded with bizarre emojis and maintaining correct spelling. In this scenario, your brain is a little bit like the man in the Chinese room: It is only with the aid of your thumbs and the smartphone that your brain can you really generate meaningful contributions to this text-message conversation.

OVERALL EVALUATION OF THE PHILOSOPHICAL APPROACH

One of the main advantages of the philosophical approach is that it allows us to ask much broader questions than those of other disciplines. A cognitive psychologist studying memory for nouns might wonder why concrete nouns are recalled better than abstract ones. This psychologist is constrained into formulating specific questions and hypotheses by the narrow focus of the research. This very focus, of course, is an advantage since it allows the researcher to scientifically examine and understand a natural phenomenon in depth. A philosopher examining the results of this same research is free to inquire about the nature of concreteness or what it means that something is abstract. He or she could also inquire as to how concrete or abstract stimuli are processed in other cognitive systems, such as attention and language. Of course, he or she is free to ask even more fundamental questions, such as, Why do we have memory? What purpose does memory serve? What would a person be like without a memory? Philosophy thus shows us the "bigger picture." It gives us key insights into the relationships between different areas of study—within and between disciplines—and, therefore, plays a very important role in the interdisciplinary endeavor of cognitive science.

Keep in mind that philosophy is a nonempirical approach. It does not use the scientific method. Concepts in philosophy are validated through logical reasoning and argument rather than by systematic observation and experimentation. For this reason, the conclusions reached in philosophy are speculative and theoretical until tested. Philosophy is better suited to the asking of important questions—how, what, and why we should study something—than to providing definitive answers. These answers come from the scientific disciplines. It is, therefore, important that a good two-way working relationship exists between philosophers and these science-based approaches.

SUMMING UP: A REVIEW OF CHAPTER 2

1. Philosophy plays a key role in cognitive science by asking critical questions.
2. According to the mind–body problem, it is not known whether mental states are physical states. Two approaches to this are (1) monism, which argues that mind and body are unitary, and (2) dualism, which states that mind and body are separate.
3. There are two versions of monism: (1) idealism, where the universe is not physical, and (2) physicalism, where it is. Identity theorists hold that the mind is the brain, while eliminativists wish to do away completely with the concept of mind as well as with any folk psychology or commonsense explanations of the mind.
4. There are also two schools of dualism. Substance dualism holds that mind and body are composed of different substances. Property dualism holds that mental states are nonphysical features of physical bodies.
5. Functionalists believe that mental states are equivalent to the functioning or operation of physical states. It may be that different material substrates can give rise to the same mental state.
6. The knowledge acquisition problem is about how mental capacities are acquired. Nativists believe that we are born with knowledge. Empiricists believe that we attain knowledge through learning and experience.
7. Consciousness is the subjective quality of experience. The "easy problem" of consciousness is explaining qualia (subjective experience) in computational or neural terms. The "hard problem" is explaining how we have such subjective experiences at all.
8. Mind may be an emergent property of a physical brain. That is, it may not be fully explained by an understanding of its component parts and then just adding those parts together.
9. The neurobiological approach to consciousness tries to discover the neural states corresponding to different types of consciousness experience, known as the neural correlates of consciousness.
10. The strong AI view is that it may be possible some day to engineer a complex system that is conscious. In contrast, the weak AI view advocates that consciousness is either nonphysical or that it is physical but too difficult to engineer.

SUGGESTED READINGS

Bechtel, W. (1988). *Philosophy of mind: An overview for cognitive science.* Hillsdale, NJ: Erlbaum.

Blackmore, S. (2012). *Consciousness: An introduction.* Oxford, England: Oxford University Press.

Chalmers, D. (1996). *The conscious mind.* Oxford, England: Oxford University Press.

Churchland, P. M. (1986). *Neurophilosophy: Toward a unified science of the mind-brain.* Cambridge: MIT Press.

Churchland, P. M. (1995). *The engine of reason, the seat of the soul: A philosophical journey into the brain.* Cambridge: MIT Press.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. New York, NY: Oxford University Press.

Dennett, D. C. (1991). *Consciousness explained.* Boston, MA: Little, Brown.

Pinker, S. (2002). *The blank slate: The modern denial of human nature*. New York, NY: Viking Press.

Searle, J. R. (1992). *The rediscovery of the mind.* Cambridge: MIT Press.

CHAPTER THREE

THE PSYCHOLOGICAL APPROACH

A Profusion of Theories

WHAT IS PSYCHOLOGY?

This chapter highlights a number of different perspectives adopted within the discipline of psychology, primarily during the 20th century. Each has a relatively unique answer to the question, What is mind? But first, we need to ask ourselves another question: What is **psychology**? As a discipline, it is best defined as the scientific study of mind and behavior. Psychology uses the scientific method as a means of gaining valid knowledge. Its focus of study includes internal mental events, such as perception, reasoning, language, and visual imagery. However, it also studies behaviors, which are external events. Behaviors include things such as walking, talking, and running. This distinction between mind and behavior makes its appearance in the fundamental questions that this approach attempts to answer. These questions include, What are the contents of the mind? How do these contents interact with one another? How does the mind explain what we do?

The different movements in psychology described in this chapter have focused on different themes. Many of the movements, including voluntarism and structuralism, were concerned with cataloging the "stuff" inside our heads. That is, they tried to list the basic elements of the mind and delineate how they interact to form other elements. We see this question of mental content taken up later by other approaches. Cognitive psychologists, who studied memory later on, tried to describe concepts that exist in our memories and how these concepts are related to one another. They formulated the idea of a mental lexicon or dictionary as a way of explaining how concepts exist in memory. In the network approach, we see that semantic networks were created explicitly to describe the location and arrangement of concepts.

A second major theme that emerges in the history of psychology is centered on operations—that is, what the mind does rather than what it contains. We see the issue of operations being addressed first

Learning Objectives

After reading this chapter, you will be able to:

1. Define what an independent variable is in an experiment.
2. Define what a dependent variable is in an experiment.
3. Discuss the difference between a theory and a hypothesis.
4. Assess the psychological method of introspection.
5. Describe the underlying principle of Gestalt psychology.
6. Critique Sigmund Freud's three components of the mind.
7. Describe behaviorism's influence on the discipline of psychology.

by the functionalists. Functionalism veered away from a listing of the mind's supposed "parts" and studied the ways in which the mind performs various mental actions. The psychoanalytic psychologists, such as Freud, were also operational in their approach. Their focus was on how mental structures such as the id and the ego dynamically interrelate. Gestalt psychology focused on a specific aspect of mental operations—namely, how mental parts form wholes. The Gestalt psychologists wanted to know how the mind creates larger-order structures during perception and problem solving. This theme of mental operations occurs repeatedly in other areas of cognitive science too. Cognitive psychologists would later form elaborate models of mental processes to explain perception, memory, and attention. In neuroscience, brain scans would reveal the neural sequences of events that underlie mental tasks. Connectionists would devise neural networks with specific patterns of operation to simulate mental function, and in the domain of artificial intelligence, computers would be programmed with complex sets of rules to mimic the kinds of things the brain can do.

Interestingly, in the behaviorist movement, we see a counterreaction to the whole concept of studying the mind. Behaviorists viewed the mind as something that passively mapped aspects of the environment onto an organism's response. It was the environment and not the mind, they believed, that controlled a person's actions. The study of behavior and its relation to the environment appears again in cognitive science in the field of robotics. A challenging goal in this field is to get a machine to perform some task successfully in a real-world environment. To do so requires an understanding of how variable stimulus inputs can map onto a robot's possible responses.

A number of commonalities run through the different movements in psychology. Many were influenced by developments in other branches of the sciences. Particularly influential were chemistry and physics. Voluntarism and structuralism both adopted fundamental ideas from chemistry and embraced the idea that the mind, like the physical world, is made up of basic elements that combine into larger wholes. Similarly, the Gestalt psychologists adopted the idea of field theory from physics and employed the notions of fields, particles, and forces in their descriptions of mental phenomena. In addition, many movements in psychology arose as counterreactions to existing movements or as a means to address problems raised by prior movements. Both functionalism and Gestalt psychology, for instance, arose in opposition to structuralism.

PSYCHOLOGY AND THE SCIENTIFIC METHOD

Before describing the foremost theories of each movement, we need to digress and examine how psychology goes about the business of testing its theories. Early psychologists relied on techniques such as introspection and phenomenology. Modern psychologists employ a wider variety of methods. They administer questionnaires and surveys, analyze case studies of single individuals, conduct psychological experiments in the lab, and record behavior in the wild through naturalistic observation. A number of modern-day psychologists also cross over into other disciplines such as neuroscience or artificial intelligence and thus employ the methodologies of these approaches as well. Many of these techniques are characterized by the principles of science.

The scientific endeavor in general is characterized by the hypothetic-deductive approach. In this approach, a hypothetical conjecture about the way the world works is tested deductively. The testing is accomplished by carefully observing the way the world works under controlled conditions. If the observation supports our conjecture, we can elaborate and expand on it. If it doesn't, we must change it to account for what has been observed. The testing, as we alluded to in the previous paragraph, can assume a variety of forms. In neuroscience, it can be the observation of brain-damaged patients or the scanning of brain activity. In artificial intelligence or cognitive psychology, it may involve constructing a computer simulation. The scientific method, in which an experiment is conducted to test a hypothesis, is perhaps the most widely used method in all of psychology and cognitive science. For this reason, we will look into it now.

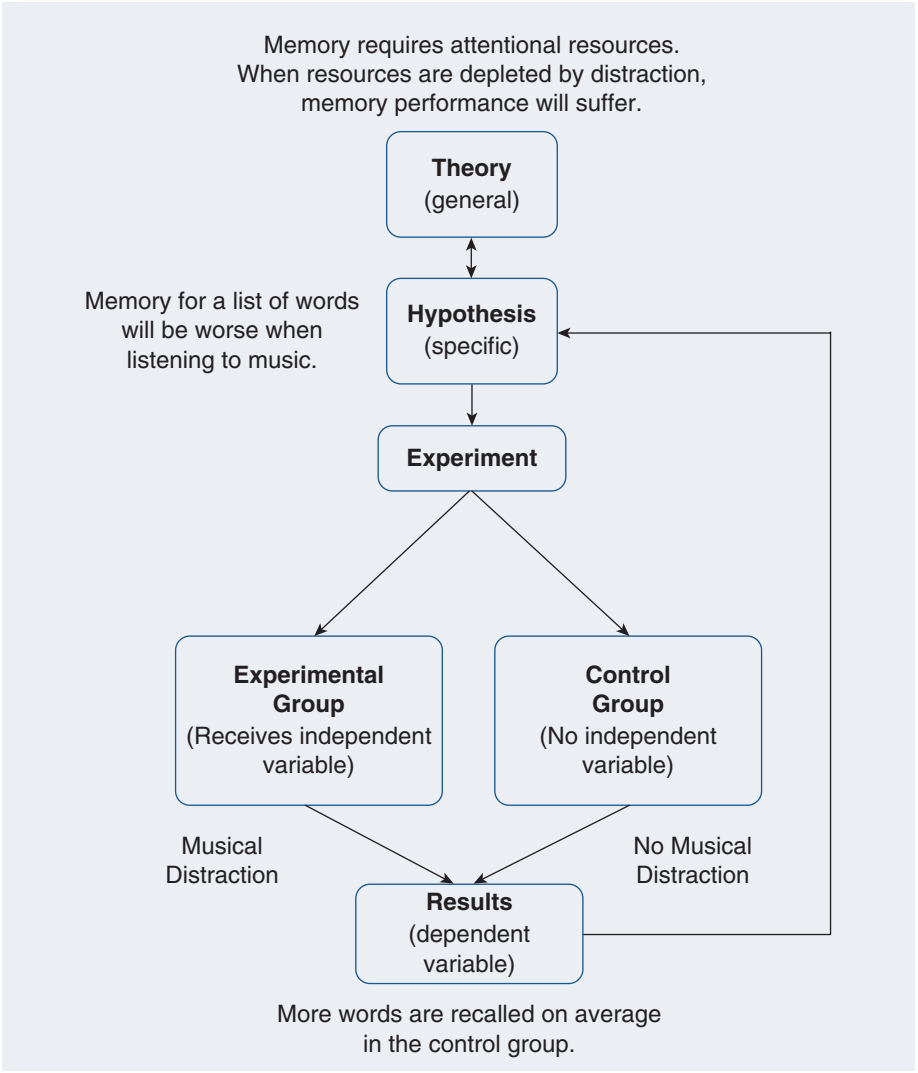
The **scientific method** uses an experiment that is designed to test a **hypothesis**—a specific statement about the world. For example, one could hypothesize that a participant's memory for a list of words will be worse when the participant is listening to music while studying them. The validity of a hypothesis is based on the outcome of an experiment. The results of an experiment can either support or fail to support a given hypothesis. Hypothesis testing helps researchers construct a **theory**—a more general understanding of the world that organizes a set of facts and aids us in understanding how the world works.

Any experiment involves the use of independent and dependent variables. An experimenter manipulates an **independent variable** to see if it will produce a change in the observed behavior. Going along with our previous example, a researcher might vary the presence of music in a memory experiment. In the music condition, participants memorize a list of 30 words while listening to music being played in the background. In the no-music condition, a different group of participants must memorize the same 30 words without any music playing. The **dependent variable** is what is measured or observed by the experimenter to see if a change of some kind has occurred. Such a change may be thought of as the effect, or the thing that happened as a result of the independent variable. In this instance, the independent variable would be music (present or absent). The dependent variable, of memory performance, could be **operationalized** in a few different ways. For example, memory performance could be measured as the average number of words correctly recalled, the speed with which correct words are recalled, or even the number of irrelevant words incorrectly recalled.

Generally, experiments consist of a minimum of two conditions. The **experimental group** is placed in one condition from the independent variable, while the **control group** is placed in the other condition. In our hypothetical example, the music condition would correspond to the experimental group and the no-music condition to the control group. Assuming that everything else that makes up the two conditions is held constant by the researchers, any difference between the two sets of results must be attributed to the manipulation. One way to help ensure that other potential factors are indeed held constant is to **randomize** the assignment of participants to the two conditions, or to **counterbalance** any known factors that could alter memory performance. For example, by randomly assigning participants to the two conditions, one could make sure that an equal number of late afternoon participants (who might be tired from a long day) wind up in each condition. Or by asking participants their age, one could counterbalance the

two conditions to have a roughly equal number of elderly participants (whose memory abilities may be reduced). Now that we've controlled potentially interfering factors, imagine we find that average word recall is higher in the no-music condition than in the music condition. Then we could conclude that background music does disrupt the memorizing of words. As stated above, this would support our hypothesis. This in turn would help us build a more general theory. The type of theory that might take shape from this and other similar experiments might be that memory processes require attention and that if attention is removed by a distractor task while memorizing, performance will suffer. The steps in this process are shown graphically in Figure 3.1.

Figure 3.1 Steps in the scientific method illustrated with an example from the text.



INTELLIGENCE TESTS

Much of this chapter focuses on theories of what the mind is made of and *how it works*. But, before we dive into that, we have one more brief digression. We spend a little time here briefly describing psychology's robust history of simply *measuring* the mind, without paying much attention to *how it works*. Intelligence tests were being developed over a century ago, when Alfred Binet was designing methods to improve France's education system. Operating on the assumption that a significant portion of a person's general intellectual capacity was innately endowed and unchanging over the life span, Binet endeavored to measure that capacity. In 1920, Lewis Terman at Stanford University adapted Binet's test for American students. The Stanford-Binet Intelligence Quotient (IQ) Test has been influential ever since.

Early on, however, there were concerns about cultural biases in the test that put disenfranchised (or less mainstream) people at a disadvantage in the test. For example, one of the early test questions for children was to name four coins (a penny, nickel, dime, and quarter). In 1920, many poor children may not have even seen a quarter yet. This would be a concrete example of how a cultural bias (in favor of high socioeconomic status) could unfairly bias the IQ score. In the 1960s, some alternative intelligence tests were developed as a demonstration of how different cultural backgrounds might disadvantage a child while taking a test designed solely by scholars from a single cultural background. For example, one test developed by African American scholars asked questions that used vocabulary terms that one might only learn in an African American cultural setting. Most African American students were able to answer the question about what Juneteenth celebrates, but many of the white students got that question wrong. Over the decades, the Stanford-Binet IQ test has undergone numerous revisions in an attempt to reduce its cultural biases.

Even with cultural biases reduced, there are still a number of critiques of IQ that aim at its core assumptions that (a) general intelligence is mostly innate, and (b) intelligence can be accurately described with a single number. Adoption studies and twin studies have shown that IQ scores can improve when a child is moved from an intellectually unsupportive environment to an intellectually supportive one (Tucker-Drob & Briley, 2014). As a result of studies like that, it is now accepted that some portion of general intelligence appears not to be innately endowed and fixed but instead can be improved during development. Moreover, others have even argued that it may be naive in the first place to assume that a single number could ever encapsulate all the different ways in which people exhibit their intelligence. Howard Gardner (2011) proposes that there are multiple aspects to what we call intelligent behavior. He identifies at least eight different types of intelligence, each of which is important for evaluating how "smart" a person can be. Writers might excel at linguistic intelligence, and musicians might excel at musical intelligence. Engineers and programmers might be strong in spatial intelligence or logical-mathematical intelligence. Talented athletes might have unusually high bodily-kinesthetic intelligence. Actors and salespersons might be exceptionally gifted in interpersonal intelligence. Others still may have strengths in intrapersonal intelligence (for self-knowledge) or in naturalistic intelligence (for a taxonomic understanding of nature). Imagining that a single IQ score could somehow capture all of those different kinds of intelligence seems impossible.

As a result of the lingering cultural biases, the flawed assumptions about genetics, and the obvious diversity of real-world intelligent abilities, the concept of intelligence in psychology has been steadily moving away from the singular notion of a genetically inherited IQ score. Recognizing that intelligence may be multifaceted and somewhat learnable, and that some people are strong in some kinds of intelligence while others are strong in other kinds, has led the field toward a more flexible notion of intelligence as something that is distributed, dynamic, and only partly determined by genetics (Turkheimer, 2000).

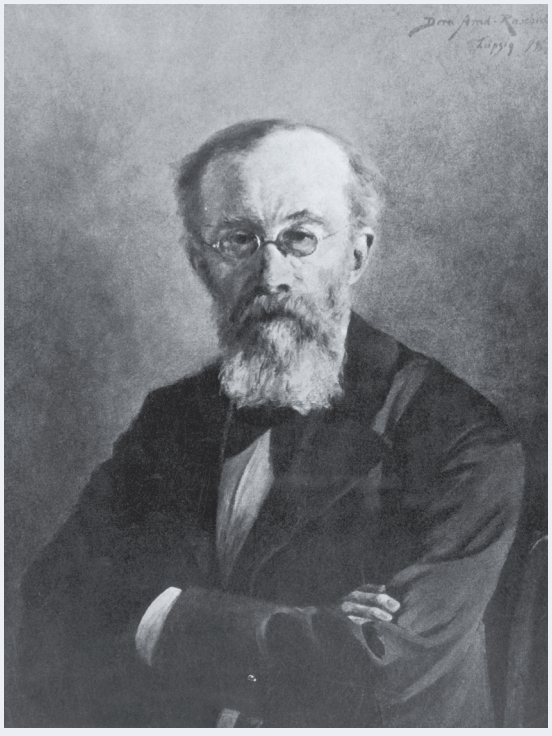
MENTAL ATOMS, MENTAL MOLECULES, AND A PERIODIC TABLE OF THE MIND: THE VOLUNTARIST MOVEMENT

Now that you've been familiarized with the scientific method, and we have carved up intelligence into a variety of mental processes, let us turn to a historical assessment of different theories for how mental processing actually works. Duane Schultz and Sydney Ellen Schultz (1987), in *A History of Modern Psychology*, give a good overview and critique of the various theoretical movements in the history of psychology. These movements include voluntarism, structuralism, functionalism, behaviorism, Gestalt psychology, and the psychodynamic view. They all preceded the cognitive revolution, which we will discuss in the next chapter. We refer our readers to their text for a more detailed discussion of the history and ideas behind these movements.

We will begin our discussion of psychology with voluntarism. The **voluntarism** movement viewed the mind as consisting of elements but stressed that these elements were assembled into higher-level cognitive components through the power of the will. It was the will, or voluntary effort of the mind, that was seen as the force behind the creation of more complex mental elements. The German physiologist and psychologist Wilhelm Wundt (1832–1920) was the founder of voluntarism (see Figure 3.2). Another scientific field, chemistry, had a significant influence on voluntarism. During Wundt's time, chemists were attempting to describe the material world in terms of its basic components and how they combine. The Russian chemist Dimitri Mendeleev developed the periodic table of chemical elements during this time. The idea behind this table was that the entire physical universe consisted of atoms—characterized by different properties—and that atoms could, under certain conditions, combine to create more complex, higher-order molecules. Wundt may also have been attempting to create a periodic table of mental elements and to specify how these elements combine (Marx & Hillix, 1979).

Wundt's method was **introspection**, or internal perception. *Introspection* literally means "inward looking." Just as one can look out at the external world to see various objects, such as a chair or table, Wundt believed that one could also look inward to experience and describe mental objects. He presented the students in his lab with various stimuli, such as colored shapes, and asked them to introspect. The students then recorded their subjective experiences about the stimuli. Although various forms of introspection had been used by philosophers for centuries, Wundt attempted to systematize and objectify the technique. He had his students put themselves in a ready state of

Figure 3.2 Wilhelm Wundt established the first experimental psychology laboratory in Leipzig, Germany, in 1879.



Source: Hulton Archive/Stringer/Getty Images.

attention prior to their introspecting and repeat their observations several times as he varied specific physical aspects of the stimulus, such as size and duration of the exposure. This kind of methodological exactness exemplifies the influence of the scientific method.

Wundt believed that psychology should study consciousness; however, he distinguished between two types of conscious experience. **Immediate experience** is our direct awareness of something. For example, if we see a rose, our perception of the rose as red is immediate. It is the redness that we experience directly while we are looking at it. If someone then asked us what we were looking at and we responded, "A red rose," then that thought would be a mediate experience (it was a thought about the rose). **Mediate experiences** are those that come from mental reflection about an object. Wundt emphasized the study of immediate experiences. He believed that they are the best way to describe the basic elements of the mind since they are "untainted" by elaborate thought processes.

Wundt went on to develop a **tridimensional theory of feeling**, according to which all feelings can be characterized by three dimensions: (1) pleasure or displeasure, (2) tension or relaxation, and (3) excitement or depression. Wundt would play a metronome, a device that produces audible clicks at different time intervals. As he varied the rate of the metronome, he discovered that some rhythms were more pleasant than others. He also reported a feeling of tension that accompanied his waiting for a click, followed by a feeling of satisfaction after the click occurred. Finally, he noted feeling excited when the tempo of clicks was increased and calm when they were played at a slower rate. Based on these introspective observations, he believed that any feeling could be characterized by pleasure or displeasure, tension or relaxation, and excitement or depression.

Although voluntarism attempted to describe and classify the basic elements of the mind, it also needed to account for the fact that we perceive unitary wholes rather than collections of individual elements. For instance, when looking at a face, we see a face in its entirety, not just a collection of its different parts: two eyes, a nose, and a mouth. To account for this, Wundt postulated the principle of **creative synthesis**, also called the law of psychic resultants. According to this principle, the mind actively organizes disparate elements together such that the resulting whole contains new properties. These new properties cannot be explained by the characteristics of the individual elements themselves. A similar phenomenon is seen in chemistry. Here, characteristics of individual atoms—for example, hydrogen and oxygen—are insufficient to specify the characteristics of the water molecules that form when they combine. Water has unique properties that cannot be understood simply by analyzing it down to its component parts.

To Wundt, creative synthesis was an active process in which the mind took hold of elements and forcefully combined them. This contrasts with the earlier views of the empiricists and associationists, who saw mental combination as a passive and mechanical effect. They argued that mental wholes were created through automatic processes of association and that these processes did not require the active participation of the mind deemed necessary by Wundt. This topic of the relationship between parts and wholes does not end here. The Gestalt psychologists would have much to say on this issue several years later.

Evaluating the Voluntarist Approach

The voluntarist approach was beneficial because it was the first scientific attempt at studying the mind. The establishment of a laboratory, the application of experimental methods, and the formulation of the clearly defined goals of listing elements and describing their combinations are all strengths of this movement. However, many of the criticisms that have been leveled at voluntarism also center on aspects of its methodology—introspection in particular. Critics have pointed to a number of flaws in introspection. First, a mental experience may change over time. A person's experience of red may undergo alteration after several seconds have elapsed—the experience perhaps becoming less vivid. Also, it may be impossible to separate immediate and mediate experiences because the act of introspecting could itself change an experience. According to this idea, the simple reporting of red, even as a sensation, involves reflective and perhaps other thought processes that alter the true experience of red. And there was the issue of

individual differences. Some observers in Wundt's lab experienced the same stimulus in different ways, which suggested that people's expectations or learning experiences could change their perceptions. Finally, Wundt was never able to compile a short list of fundamental mental elements comparable to the list of elements that had been compiled in physical chemistry. The list of elements derived from his students' introspections with respect to their immediate, sensory-type experiences was growing far longer than what a simple periodic table would accommodate.

STRUCTURALISM: WHAT THE MIND IS

Structuralism shares a number of ideas with its predecessor, voluntarism. The subject matter of psychology was again conscious experience, and the method was a modified version of introspection with an emphasis on the scientific method. There were, however, major theoretical differences between the two schools. The structuralist view of the mind held, once again, that the mind was a passive agent, with mental elements combining according to mechanistic laws. Structuralism, per its name, focuses on mental elements; that is, the structure of the mind is to be understood in terms of basic elements and their combination—suggesting once again the analogy to chemistry. Its progenitor was the U.S. psychologist Edward Bradfort Titchener (1867–1927).

Titchener cautioned against making the **stimulus error**, which is to say, confusing our true experience of an object with a description of the object based on language and past experience. When one looks at a banana, the true experience would be that it is yellow and curved—not the recognizing of it as a banana or a type of fruit. This distinction parallels Wundt's differentiation between immediate and mediate experience. Titchener, in addition, broke away from Wundtian-style introspection. He believed that only well-trained observers could introspect accurately and not make the stimulus error.

According to Titchener, psychology had three goals: (1) to describe consciousness in terms of its simplest and most basic components, (2) to discover the laws by which these elements or components associate, and (3) to understand the relation between the elements and their physiological conditions. The initial aim was, as with voluntarism, to come up with a set of fundamental mental units. Titchener believed that a mental element was fundamental when it remained constant over numerous introspective trials. So if multiple trained introspective observers all consistently experienced an object in the same way, then this experience qualified as a true element.

The combination of elements, in Titchener's scheme, was not affected through active mental processes, as Wundt believed. Instead, Titchener saw the mind as a passive mechanism or substrate within which elements combined according to set laws. In chemistry, a **reagent** is a substance added to a mixture to produce a particular chemical reaction. Reagents are used to quantify these reactions and their chemical products. In a similar fashion, structuralists believed the mind to be a reagent, a medium inside of which mental reactions and processes unfolded. Researchers of the day even went so far as to refer to their experimental participants as reagents.

Like Wundt, Titchener listed his elements. He described a total of 44,000 sensation elements alone. Out of these, 32,820 were visual and 11,600 were auditory

(Titchener, 1896). He thought that each of these elements was fundamental and indivisible and that each was capable of combining with others to form more complicated perceptions and ideas. Titchener believed that all sensations could be characterized by four attributes. **Quality** is what separates a sensation from any other sensation. The experience of heat, therefore, has a different quality than that of sound. **Intensity** refers to how strong a sensation is; for example, a noise can be loud or soft and a light can be bright or dim. **Duration** refers to how long a sensation persists, whether it is short-lived or longer lasting. Sensations are characterized also by **clearness**. Sensations that one pays attention to possess greater clarity. To these basic four attributes, Titchener later added **extensity**, the extent to which a sensation fills or occupies space. The sensation of pressure coming from a pencil tip that touches the skin has less extent than that coming from a chair bottom.

Evaluating the Structuralist Approach

Many of the advantages and disadvantages of the structuralist movement are identical to those of its ancestor, voluntarism. Structuralism further refined scientific methodological procedures and applied them to the study of psychological phenomena. It employed a variety of techniques, including measurement, observation, and experimentation. Introspection in a modified form would be put to use later in other areas of psychology, such as **psychophysics**—in which observers judge the relative intensity of stimuli—and clinical psychology—in which participants are asked to evaluate themselves along personality continua, such as those centering on anxiety or depression (Schultz & Schultz, 1987).

More elaborate critiques of introspectionism were now pointed at the structuralist school. Among these were that the process itself is subjective and unreliable, that training participants to introspect only biases their responding even more, and that some mental experiences, such as habits, happen without conscious awareness and so cannot be introspected about at all. The psychoanalytic school would expound considerably on this idea of an unconscious mind that is inaccessible to attention or awareness. In conclusion, structuralism was found to be too analytic and reductionistic. It was found to overemphasize the role of low-level mental elements and to ignore holistic perception and experience, once again opening the way for the Gestalt psychologists' concern with this issue.

FUNCTIONALISM: WHAT THE MIND DOES

In the scientific disciplines, a theoretical perspective that has been entrenched for many years is often quickly replaced with an alternative, sometimes completely contrary, perspective. This was certainly the case with functionalism's supplanting of structuralism. Rather than focusing on the contents of the mind, functionalism focused on what the mind could do. Its emphasis was on the mental processes or functions that operate on the elements, instead of the elements themselves. Harvey A. Carr (1873–1954), one of the later U.S. functionalists, summarizes the subject matter of functionalist psychology as follows:

> Psychology is primarily concerned with the study of mental activity. This term is the generic name for such activities as perception, memory, imagination, reasoning, feeling, judgment, and will. . . . Stated in comprehensive terms, we may say that mental activity is concerned with the acquisition, fixation, retention, organization, and evaluation of experiences and their subsequent utilization in the guidance of conduct. (Carr, 1925, p. 1)

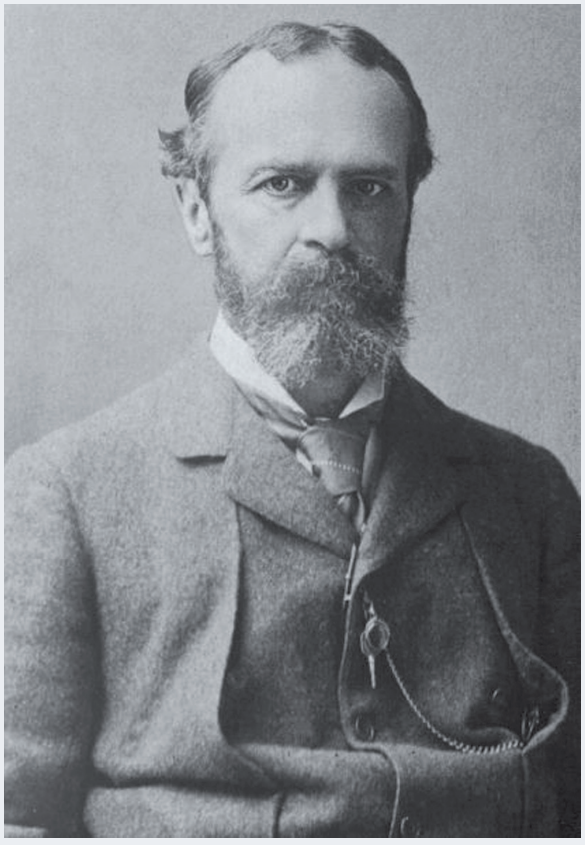
The formal development of psychological functionalism in the United States is credited to John Dewey (1859–1952) and James Rowland Angell (1869–1949), as well as to Harvey A. Carr. However, it was William James (1842–1910) who was its pioneer and perhaps its most lucid expositor (see Figure 3.3). James is often considered the greatest American psychologist. James rebelled against the Wundtian and structuralist conception of experience as being made up of discrete elements and believed that such elements existed simply as a result of the attentive, introspective process. He thought that an individual erroneously created the notion of an element by attempting to "freeze" or stop a moment in the mind's ongoing activity. He stated that one person's subjective element in response to a perception, for instance, does not guarantee that the same element will exist in anyone else's mind who experiences the same perception—what he called the **psychologist's fallacy**.

James replaced the "mind-as-elements" notion with the idea of the mind as a **stream of consciousness**, suggesting that the mind was a process undergoing continuous flow or change. He likened thought to the water in a river that is always moving. He provided another analogy when, referring to mind, he declared, "Like a bird's life, it seems to be made of an alternation of flights and perchings" (James, 1890, p. 243). James termed the resting places or "perchings" of thought as *substantive*. **Substantive thought** occurs when the mind slows down, perhaps when focusing attention. James called the "flights" of the mind **transitive thought**. They correspond to a less focused, more associative form of thinking.

James Rowland Angell (1907) articulated the three major themes of functionalism. The first theme was the study of **mental operations**. By this, he meant that functionalism should investigate how a mental process operates, what it accomplishes, and under what conditions it occurs. Second, Angell believed that functionalism should study the **fundamental utilities of consciousness**. One ought to understand the role consciousness plays in the survival of the organism. This means asking questions such as, Why are we conscious? Of what utility or usefulness is consciousness? How does consciousness help keep an organism alive? Third, functionalism should study **psychophysical relations**, the relations between the psychological mind and the physical body. The term also refers to the total relationship of the organism to the environment and to how the organism, both physically and mentally, may have adapted to the environment.

From these themes, it is clear that functionalism was strongly influenced by Darwin's theory of natural selection. Functionalism then sought to explain not just mental operations in and of themselves but also how the mind in general arose through evolutionary pressures to serve the organism. The idea of an organism–environment fit

Figure 3.3 William James was one of the early American psychologists.



Source: FPG/Staff/Archive Photos/Getty Images.

and how it may help explain a variety of different psychological characteristics was to be elaborated on more fully at a later date by evolutionary psychologists. Functionalism thus serves as the theoretical precursor to evolutionary psychology, discussed in greater detail later in this book.

Evaluating the Functionalist Approach

The functionalist movement broadened the field of what could acceptably be studied in psychology. Unconscious phenomena were fair game, as was the study of children, the mentally disabled, and the "insane." It allowed for a wider variety of methods, such as

tests, questionnaires, and objective behavioral descriptions, and it permitted the continued use of introspectionism (Schultz & Schultz, 1987).

Criticisms of functionalism came, perhaps not surprisingly, from the structuralist camp. In 1913, C. A. Ruckmick, a student of Titchener's, accused the functionalist movement of failing to define the term *function*, having noted that functionalists used the term in two different ways. First, *function* was used to refer to an activity or process itself, such as perception or memory. Second, *function* was used to mean the utility or usefulness of an activity to the organism (Schultz & Schultz, 1987). Functionalists sometimes used these terms interchangeably, which occasionally invited confusion. Another reported fault of functionalism was that it was too applied and practical. Its adherents were accused of focusing excessively on the usefulness of functions. In contrast, the structuralists took a more basic scientific approach in their attempt to describe and elucidate the basic aspects of the mind. This debate over which is better—basic or applied research—is ongoing in psychology and other sciences.

THE WHOLE IS GREATER THAN THE SUM OF ITS PARTS: MENTAL PHYSICS AND THE GESTALT MOVEMENT

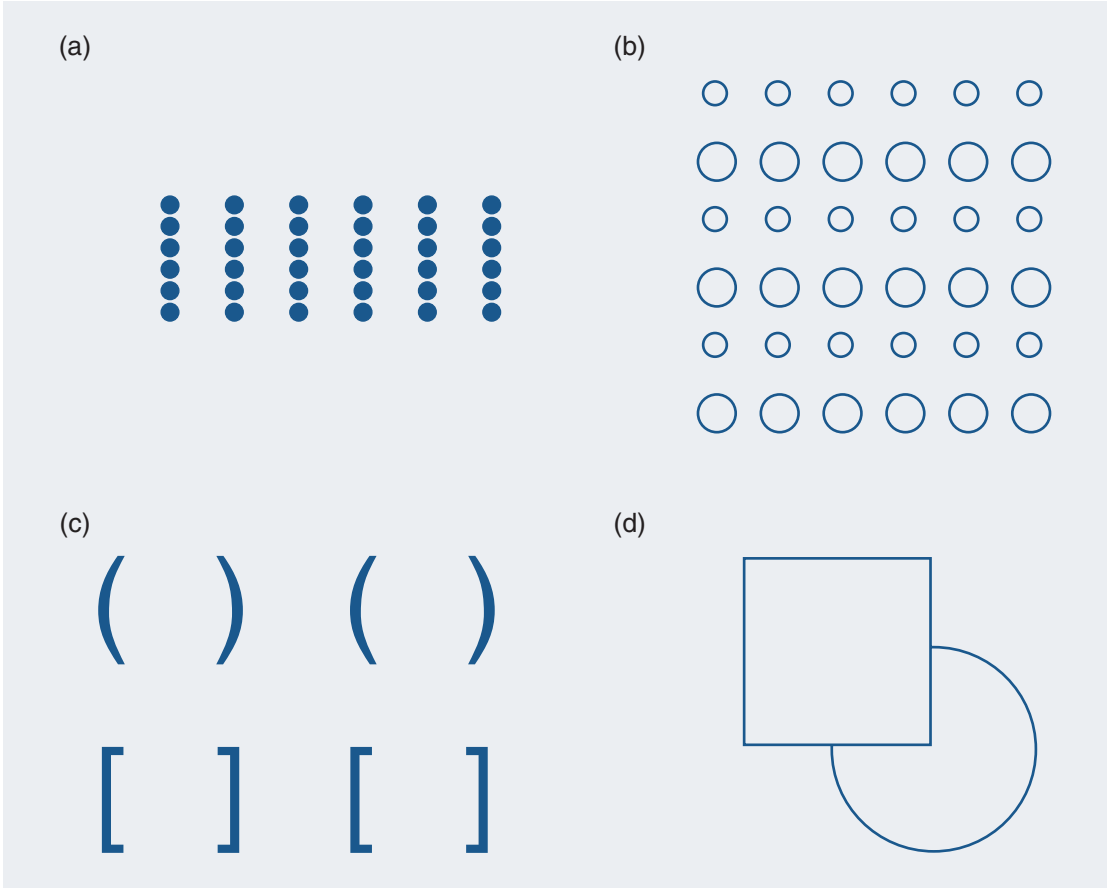
In the early part of the 20th century, another theoretical framework arose in psychology. This framework was called Gestalt psychology, and its three founders were all German. They were Max Wertheimer (1880–1943), Kurt Koffka (1886–1941), and Wolfgang Kohler (1887–1967). The Gestalt movement, like functionalism, was a counterreaction to structuralism and the atomism it entailed. The Gestalt psychologists even referred to structuralism as "brick-and-mortar psychology." They saw wholes as more than just the sum of their parts; such an integrated whole they called a **gestalt**. What characterizes the Gestalt approach is the importance of conscious wholes. The Gestalt psychologists believed that conscious wholes could not simply be reduced to a listing and description of their parts. The Gestalt psychologists also borrowed a metaphor from physics: They believed that mental parts combined into wholes in much the same way that physical particles organized when subjected to fields of force.

The contributions of Gestalt psychology were greatest in two areas: (1) perception and (2) learning. Gestalt psychology is phenomenological in method. **Phenomenology** refers to subjective experience rather than objective description. When studying perception, for example, the Gestaltists preferred to create stimulus patterns, show them to observers, and have the observers describe their subjective experiences. Phenomenological description differs from introspection in that it focuses on a person's immediate and subjective perception of an external stimulus. It does not require training or an intensive examination of one's internal state. The Gestalt psychologists were looser methodologically when studying learning as well. They preferred to observe human or animal subjects finding solutions to problems, rather than setting up rigorously controlled experimental conditions.

In visual perception, the relationship between parts and wholes is an important one. As mentioned earlier, when we look at an object, we tend to see the entire object rather

than the parts of which it is composed. We tend to see a face, not an aggregate of eyes, nose, and mouth. We tend to see trees and not collections of leaves and branches. Max Wertheimer (1923) illustrates the problem this way: "I stand at the window and see a house, trees, sky. Theoretically I might say there were 327 brightnesses and nuances of color. Do I have 327? No. I have sky, house, and trees" (p. 301).

Figure 3.4 **Dot lattices and other figures that demonstrate several of the Gestalt principles of perceptual organization. In the lattice in (a), the organization is seen in terms of columns, not rows, because the dots in that orientation group by proximity. In (b), the organization is one of rows because the dots group by similarity of size. The inward-pointing pairs of parentheses and bracket-like figures in (c) form a coherent whole because of closure. The shapes in (d) are almost always perceived as a square overlapping a circle, rather than as a Pac-Man and a square, because a circle is simpler according to the law of pragnanz.**



Wertheimer then went on to formulate the **principles of perceptual organization**. These are ways in which visual parts group to form objects. The principles demonstrate that the relationships between the parts are important in determining how these parts are assembled into wholes. According to the principle of **proximity**, parts that are close to one another in the visual field tend to be perceived as connected and part of a whole. Here, the physical distance between elements is a relationship independent of those elements but one that serves to group them together. According to the principle of **similarity**, parts that are similar in lightness, color, shape, or some other characteristic also tend to be perceived as grouped together. The relationship here consists of the elements' shared properties. The principle of **closure** holds that the parts that form a complete (or almost complete) and enclosed object go together. Wertheimer also proposed the idea of **pragnanz**, or "good figure." According to pragnanz, parts that are simple will group together. Simple organizations are those that have fewer parts or are bilaterally symmetrical. Wertheimer created **lattices**, or matrices of regularly arranged parts, to demonstrate these principles. Several lattices and other Gestalt stimulus patterns are shown in Figure 3.4. It should be noted that these are only a few representative examples of grouping principles proposed by the Gestaltists. More recent perceptual investigators have suggested others.

INTERDISCIPLINARY CROSSROADS: GESTALT PHENOMENOLOGY, EXPERIMENTAL PSYCHOLOGY, AND PERCEPTUAL GROUPING

As mentioned earlier, the Gestalt psychologists, such as Wertheimer, formulated the principles of perceptual grouping to explain how parts are grouped together to form wholes. This type of process is necessary for us to see objects rather than bits and pieces of the visual environment. However, these principles as originally formulated are qualitative. They tell us why some things go together but don't tell us what the strength of grouping is. In addition, they don't tell us what grouping is like when two or more principles are at work. Remember also that these rules were created on the basis of phenomenological evidence where observers looked at patterns and verbally reported what they saw. Verification of these observations is needed using the scientific methods of modern-day experimental psychology.

Recent work has gone some way toward achieving these goals. Contemporary researchers have turned once again toward using dot lattice stimuli to estimate perceptual grouping. It turns out that there are a total of six different lattice types: (1) square, (2) rectangular, (3) centered rectangular, (4) rhombic, (5) oblique, and (6) hexagonal (Bravais, 1949; Kubovy & Wagemans, 1995). Examples of each are shown in Figure 3.5. The distances between points in different directions can be varied, as can the similarity of the elements in relation to one

(Continued)

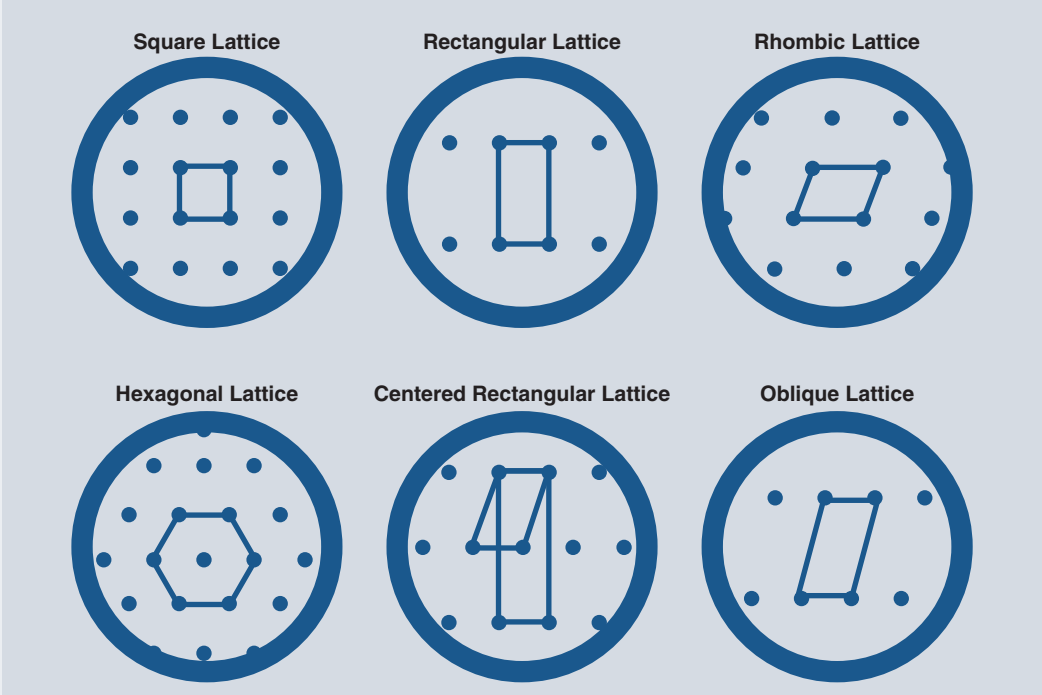
(Continued)

another. For instance, we can make the dots in one direction closer to one another, causing them to group by proximity. At the same time, we can make the dots in that direction or in other competing directions more or less similar based on various types of similarity, such as color, size, or luminance.

Using these stimuli, Kubovy and Wagemans (1995) obtained an attraction function that expresses the probability of grouping over distance. The function was an exponential decay, giving a quantitative measure of how grouping strength weakens with increased separation between dots. A later reanalysis of these data showed that this grouping relationship remained constant with changes in scale (size) or time (Kubovy & Holcombe, 1998). In another study, Lars and Kubovy (2006) employed lattices where one organization was linear (i.e., the dots in one direction formed a line), while the rest were curvilinear, forming curves. In the competition between these two types of organizations, the curved configuration was stronger. What can account for the saliency of curvature? The authors speculate that a smoothly curved alignment of dots is less likely to have occurred by accident and thus more likely to be the result of a structured object than are straight lines (see also Field, Hayes, & Hess, 1993). Curved lines are more prevalent in the natural world, so it is possible that our brains may have evolved or adapted to be more sensitive to the presence of curvature.

But what about other cases where two or more grouping principles are pitted against one another? A gestalt is often defined as a whole that is greater than the sum of its parts. This

Figure 3.5 Examples of the six different dot lattice types.



suggests that if two principles such as proximity and similarity both predict the same organization, their combined effect should be stronger than both of their individual strengths together. One might predict, for example, that their effects are multiplicative and equal to the multiplication of their separate forces. It turns out that this is not the case. Perceived grouping in dot lattices is actually additive (Kubovy & Van den Berg, 2008).

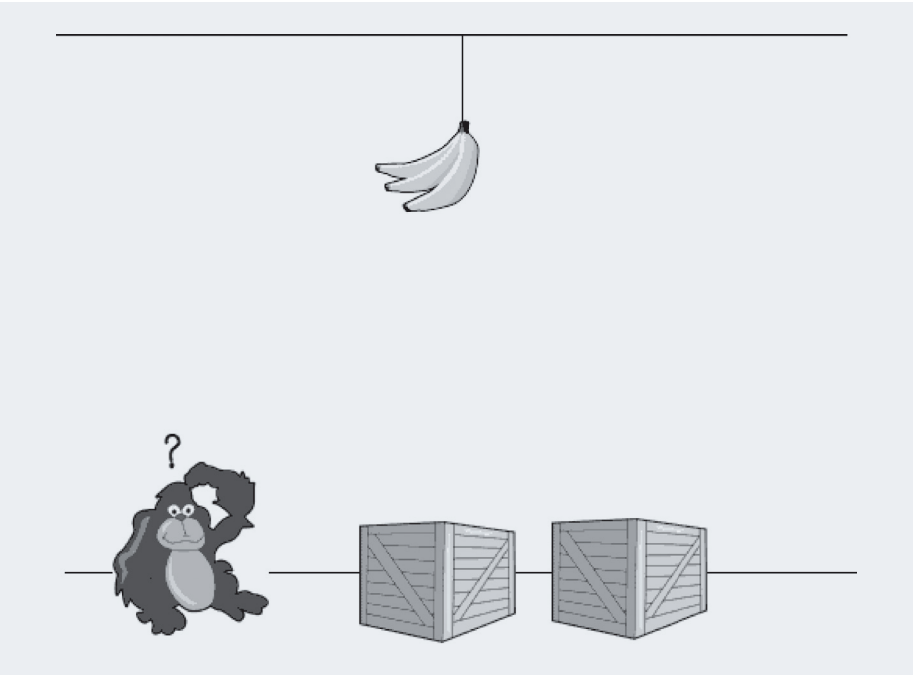
These studies show that it is possible to take an old approach and combine it successfully with a new one. Phenomenology as employed by the earlier Gestalt psychologists is qualitative and nonempirical. It is useful for demonstrating basic principles or forming hypotheses. But these ideas must be submitted to the rigor of modern experimental methods if we are to validate any of their assumptions. The newer research described here may be thought of as phenomenological psychophysics because it relies on subjective report while utilizing the methodology of psychophysics (Kubovy & Gepshtein, 2003). Psychophysics is the use of quantitative methods to measure relationships between stimuli, which are physical, and perception, which is psychological.

Wolfgang Kohler's greatest contribution was in the area of animal learning. Kohler (1927) studied chimpanzees on the island of Tenerife during World War I and described those studies in *The Mentality of Apes*. He would place a banana on a ceiling hook and observe how the chimps would use available materials to try to reach it. The chimps at first would use different approaches at random—trying to knock the banana down with a stick, for example. He then noticed that the chimps, with great suddenness, often after a period of inactivity, would solve the problem by stacking crates on top of each other and climbing that stack of crates to reach the prize (see Figure 3.6). Kohler termed this behavior **insight learning**, the ostensibly spontaneous understanding of relationships that produces a solution to a problem. He believed that this phenomenon was operational in humans as well. Because the solution is a holistic configuration and involves a set of interrelationships among component parts, it demonstrates again the Gestaltists' emphasis on mental wholes. These wholes can be perceptual, as is the case during the perception of an object, or conceptual, as is the case during learning. It should be stressed that these wholes in the Gestalt view consist not just of individual parts but, additionally, of the relationships between them.

According to Wallas (1926), insight learning happens in four stages. The first is **preparation** and consists of the acquisition and understanding of the problem as well as preliminary attempts at solving it. The second stage is **incubation**, whereby the problem is put aside for a while. During this period, there are no conscious attempts at problem solving, but the unconscious mind may be attempting to find or may have found a solution. In the third stage, **illumination** occurs. It is a flash of insight, a sort of "Aha!" experience in which the solution comes suddenly to awareness. Finally, there is **verification**, in which the insight is confirmed and one checks to see that it yields a correct solution.

The Gestalt psychologists formulated a theory of how underlying brain mechanisms give rise to Gestalts. Wertheimer suggested that the cerebral cortex was a dynamic system, in which the elements in force at a given time interacted. Kohler (1920), in *Static and Stationary Physical Gestalts*, further proposed that cortical processes operate as do the fields of force that are studied in physics. Metal filings caught in an electromagnetic

Figure 3.6 If you were the chimpanzee, how would you reach the bananas?



field generated by the poles of a magnet will organize themselves around the poles. These and related phenomena in physics are described by **field theory**. Similarly, it was believed that perceptual elements spontaneously organized themselves because they were caught in "mental force fields." In this mental equivalent of field theory, sensory impulses coming from a stimulus initiate neuronal activity. This electrochemical activity then produces a mental field that causes the perceptual grouping. Here again, we see the influence on psychology of another discipline in the natural sciences.

The Gestalt psychologists believed that there was a correspondence between the psychological or conscious experience, on the one hand, and the underlying brain experience, on the other. This perspective is called isomorphism because the perception is identical (Greek: *iso*) in form or shape (Greek: *morph*) to the brain activity that gives rise to it. The mental representation of elements and the fields operating on them were thought to be quite similar to the conscious perception of the stimulus, in the same way that a map is similar to the geography it represents.

Evaluating the Gestalt Approach

Gestalt theory gave an interesting alternative formulation of the part–whole problem. It addressed the problem in a more detailed fashion than had the

structuralists or voluntarists. It also spawned research traditions in perception and in learning and problem solving. However, critics soon pointed out a number of flaws. The Gestalt psychologists' methodology was the first aspect of Gestalt theory to receive harsh treatment. The phenomenological approach was labeled as "soft" and lacking in scientific rigor. Gestalt psychologists were accused also of being too theoretical and of failing to back up their assertions with empirically derived data. Much of their data, because these were not obtained within controlled experimental settings, were not amenable to statistical analysis. Their approach was thus critiqued as too qualitative and not quantitative enough in orientation. Field theory was additionally accused of being speculative and based on poorly defined physiological assumptions.

Some of the specific findings of the Gestalt researchers were also attacked. The principles of perceptual organization were said to be merely descriptive; they described how parts grouped but did not provide a real explanation. Field theory as an explication for grouping was seen as too vague as well as inadequate. Critics averred that the organizational laws seemed sufficient to describe grouping when one principle was at work but failed to make predictions about organization taking place under more complex real-world circumstances. Figure 3.7 shows three dots—A, B, and C. Dots A and B group by proximity, while Dots B and C group by similarity. What is the resulting organization? A further criticism was that the concept of pragnanz was ill defined. Stimuli can be constructed where a more complicated, not simpler, organization is perceived. Gestalt findings in learning were also not immune to such evaluations. Kohler's idea of insight learning as something that happened quite rapidly was questioned. Later investigators found that insight learning in animals such as chimpanzees does not always occur suddenly and can depend on factors such as prior experience and learning (Windholz & Lamal, 1985).

Figure 3.7 **The first two dots group by proximity. The second and third dots group by size similarity. The resulting perceptual organization cannot be predicted by the individual laws of organization alone.**



MINI MINDS: MECHANISM AND PSYCHOANALYTIC PSYCHOLOGY

Psychoanalytic psychology as promulgated by Sigmund Freud (1856–1939) saw the mind as being made up of distinct components, or "miniature minds." Each of these

minds competes with the others and vies for control of behavior. Psychoanalytic psychology posited not just one state of consciousness but three and emphasized the role of the unconscious mind, of which the individual has little awareness and over which he or she has little control in influencing thought and action. Freud also thought that sex, pleasure, aggression, and other primitive motivations and emotions were powerful influences on personality, as were early childhood experiences.

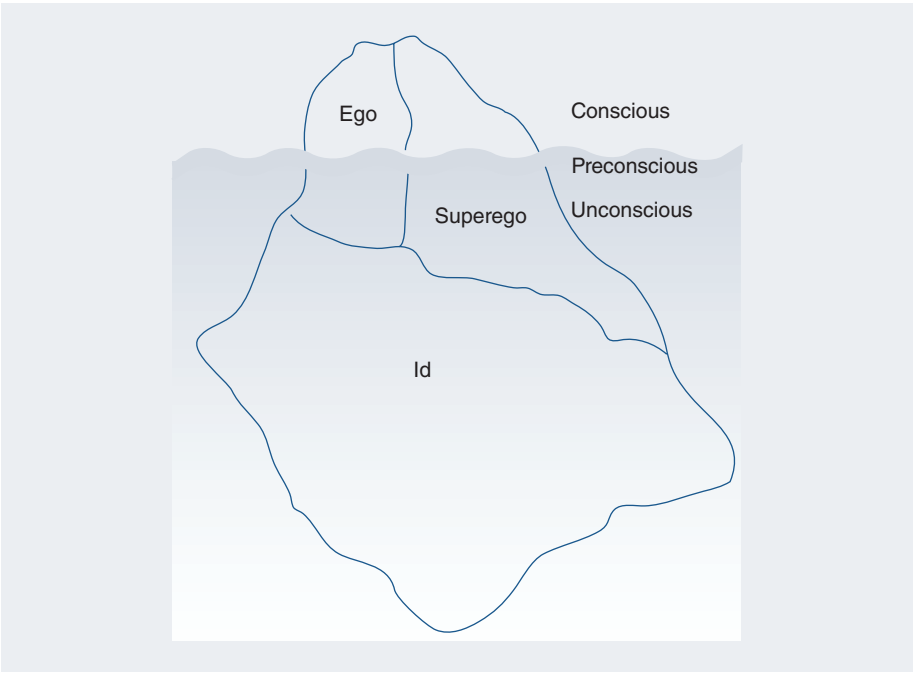
Freud proposed a three-tiered system of consciousness: (1) the **conscious** mind, which contains those thoughts and feelings of which we are aware and can directly access; (2) the **preconscious** mind—that is, those aspects of the mind that we can bring into awareness with effort; and (3) the **unconscious** mind, or the aspects of the mind of which we are completely unaware. An example of conscious content would be knowing the street address where one lives. Preconscious content requires some mental effort for one to be made aware of it. An example would be trying to recall and eventually recalling what one did on his or her birthday the past year. Unconscious content may never come into awareness, although one goal of psychoanalytic therapy is to try to allow this to happen. An example of unconscious content is memories of childhood abuse or other traumatic experiences from the early developmental years.

Freud described three other mental structures, each having a different operating principle. The **id** contains unconscious impulses and desires such as sex and hunger. It operates on the **pleasure principle** and attempts to attain gratification for its desires immediately. The **superego** is responsible for our ethical sense. It operates on the **idealistic principle** and motivates the individual to do what it considers morally sound or proper. The **ego** balances the competing demands of the id and superego. It operates on the **reality principle**, spurring one to act in a rational and pragmatic fashion.

A metaphor Freud proposed for thinking about these structures was an iceberg, depicted in Figure 3.8. He likened the conscious mind to the visible part of an iceberg above the waterline; thus, it is always visible. The preconscious mind is equivalent to the part of the iceberg near the waterline since it is sometimes submerged and unconscious but sometimes above the waterline and inside awareness. The largest part of the iceberg is then the unconscious, which is always submerged and hidden from view. The id lies completely in the unconscious, its large mass indicating its power. Both the ego and the superego occupy all three states of consciousness.

A crucial part of this perspective is the dynamic interplay between the id, ego, and superego. The id demands immediate satisfaction of its appetite for pleasure, while the superego in most cases advocates controlling or suppressing such urges in the name of decency. It then falls to the ego to attempt a rational and viable solution to the conflict. The ego is thus being barraged with the id's impulses and the superego's restraints. If it fails to satisfy either one, the result is anxiety. To shield itself against anxiety, Freud believed that the ego constructs **defense mechanisms**, which reduce or redirect anxiety in various ways. Examples of defense mechanisms include **repression**, the banishing of anxiety-arousing thoughts and feelings from consciousness, and **sublimation**, the transformation of unacceptable impulses into socially valued motivations. If a person lost memory of his or her being abused as a child, Freud would say that he or she repressed the memory because it was too painful. If instead that person channeled the anger he or she felt as a result of having been abused toward becoming successful in a career, he or she would be sublimating.

Figure 3.8 Sigmund Freud's iceberg model of the mind.



Mechanism and determinism are inherent in the psychodynamic view. One can envision Freud's model of the mind as an elaborate machine with different interacting parts. In his writings, he uses many terms drawn from mechanics, electrical circuitry, and hydraulics. Without too much imagination, it is possible to construct a Freudian hydraulic or "plumbing" model of the mind, with pumping stations connected by valves and tubes. Id impulses in this model could be represented as an increase in water pressure. Forces deriving from other components that release or redirect this pressure could counteract that increase. This view of the mind as a machine implies determinism as well. Freud thought that all mental events, including slips of the tongue and dreams, were determined and that nothing in human thought or behavior could occur by chance.

Evaluating the Psychoanalytic Approach

Many of the ideas in the psychoanalytic approach have permeated the wider culture. Freud drew our attention to important issues such as the unconscious and the influence of biological forces. The psychoanalytic approach served to stimulate further research in these areas and inspired several generations of clinical practice, but its lasting legacy may be in the strong counterarguments it elicited.

The neo-Freudians, although they accepted most of the substantive parts of Freud's theory, differed in several respects. They placed more emphasis on the role of the conscious mind and downplayed the importance of sex and aggression as all-powerful motivators. Modern psychological research suggests that Freud overestimated parental and early childhood influence and underestimated the importance of other social factors, such as peer influence (Frieze, Parsons, Johnson, Ruble, & Zellman, 1978). In addition, repression has been found to be a relatively rare mental response to psychological trauma. In fact, it is typically the case that horrible memories are remembered quite clearly (Loftus, 1995). The unconscious mind is now viewed not so much as a seething cauldron of primitive urges but as a reservoir of information processing that goes on without awareness (Greenwald, 1992).

Psychoanalytic psychology also has been critiqued for its scientific shortcomings. The theory is not based on objective observations. Its raw material comes from Freud's subjective notes about his patients, written several hours after therapy. Freud additionally offers few hypotheses that can be scientifically verified or rejected. Some have criticized psychoanalytic descriptions as being incapable of being proved one way or another. For example, the presence of anxiety in an individual could indicate the fear of giving in to id impulses. The absence of such anxiety could indicate its repression. The perspective thus offers after-the-fact explanations for personality. It also fails to predict behavior or traits (Myers, 2001). One could in this framework explain why someone is angry but not anticipate accurately when he or she might get angry in the future.

MIND AS A BLACK BOX: THE BEHAVIORIST APPROACH

The weaknesses of the psychoanalytic position were the behaviorists' strengths. Instead of attempting to describe complex mental operations, the behaviorists, as their name implies, focused entirely on the study of behavior. In their view, the internal workings of the mind were simply too complex as well as incapable of being measured in an objective and scientific fashion. Behaviors, however—the actions that are produced by organisms, such as running, grasping, or bar pressing—are external, which makes them easily measurable and quantifiable. They are thus ideally suited for scientific study. Behaviorism was influenced by work in animal and comparative psychology. In these disciplines, there was a trend to discover which environmental conditions might cause animals to act in certain ways, as opposed to an emphasis on constructing elaborate mental models of how they might be thinking. Humans were de facto lumped into the same category as animals, making behaviorism a more general science and an extension of the more rigorous natural sciences.

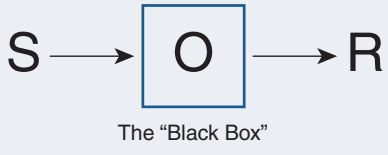
The behaviorists clearly saw themselves as true scientists. Because the mind could not be studied scientifically during their time, they preferred not to study it at all! This did not mean that behaviorists, such as B. F. Skinner, denied the existence of the mind or that they thought that brains were unworthy of study. Rather, they believed that the scientific method could not be appropriately applied to the mind and so redirected their empirical investigations to something that was amenable to the scientific method, namely, behaviors.

Behaviorism did, however, reject outright the study of consciousness through introspection, its adherents claiming that the goal of psychology was not to study consciousness at all but, instead, behaviors and the antecedent conditions that give rise to them. Figure 3.9 shows a version of the stimulus–response model that represents their position. In the model, a stimulus (S) in the environment impinges on an organism (O). The stimulus then causes the organism to produce a response (R). The most important class of stimuli that influences responses, according to behaviorists, is rewards and punishments. In this case, a stimulus might consist of a reward given to an organism—for example, a rat—which would then cause the rat to respond by continuing to press a bar. Notice that in this model, the mind, contained in the organism, is not depicted. It is treated as a "black box," meaning that it is unfathomable and unexplainable. The mind becomes simply an entity that serves to convert stimuli to responses.

Important figures associated with the movement were Edward Lee Thorndike (1874–1949), Ivan Pavlov (1849–1936), and Burrhus Frederick Skinner (1904–1990). Pavlov was a Russian physiologist who studied the dog's digestive system. In his lab, he noticed that dogs would salivate not only when eating but also immediately before. This was true only for dogs that had been fed in the lab for some time. He assumed that, in addition to the food itself, any stimulus that was put forward repeatedly just before the food was presented would also be able to trigger salivation. To test this idea, Pavlov applied meat powder to the dogs' tongues. This caused them to salivate. Then, he sounded a tuning fork each time, just prior to administering the powder. He did this over several trials. Finally, he sounded the fork without presenting any powder and found that the dogs still salivated.

As a result of these experiments, Pavlov was inspired to propose several terms. The **unconditioned stimulus** is that which elicits a response all on its own, before any training. In this instance, it is the meat powder. An **unconditioned response** is a response elicited by an unconditioned stimulus—here, the salivation caused by the administration of the meat powder. A **conditioned stimulus** elicits a response only after it has been paired repeatedly with an unconditioned stimulus—here, the sound of the tuning fork. Last, a **conditioned response** is a response elicited by a conditioned stimulus—here, the

Figure 3.9 The behaviorist stimulus–response model of behavior. The mind of an organism (O) that emits a response (R) to a stimulus (S) in this view could not be well understood. The mind was treated as a "black box."



salivation caused by the tuning fork. This form of learning came to be known as **classical conditioning**. Note that it pertains only to built-in reflexes and does not require any conscious thought on the part of the animal. The dog learns simply by associating the two stimuli.

Skinner was responsible for uncovering more about the nature of learning. Whereas classical conditioning works only with involuntary reflexive behaviors, such as the salivatory response, operant conditioning is pertinent to any voluntary motor act and is thus far more general. **Operant conditioning** is that type of learning in which a behavior is strengthened if it is followed by reinforcement and diminished if followed by punishment. A **reinforcement** is anything that increases the frequency of a response. Food is a good reinforcer for most animals. A child, if given a cookie for taking out the garbage, will be more likely to take out the garbage in the future. **Punishment**, on the other hand, is any consequence of a behavior that is apt to decrease the future incidence of that behavior. A painful stimulus such as electric shock can serve as an effective punishment. Lab rats repeatedly shocked for touching an object in their cage will soon cease to do so. Skinner was able to use various forms of reinforcement and punishment to control the behavior of animals such as rats and pigeons.

Evaluating the Behaviorist Approach

The strength of the behaviorist position is its adoption of a completely objective science of behavior. Behaviorists understood that mentalistic concepts were poorly defined and could be measured only via subjective methods such as introspection. They were very rigorous in their application of the scientific method to the study of psychological phenomena. As such, they were confident that their results yielded the most accurate and valid information. The testimony to behaviorism's success is that it was the dominant paradigm in research psychology for approximately 50 years. It didn't begin to decline until it was challenged by the rise of cognitive psychology in the 1960s. Some of the specific challenges posed by the cognitive movement will be discussed in the next chapter.

Behaviorism had allowed for the controlled and systematic study of behavior, but the behaviorists' abandonment of consciousness and mind was the movement's greatest weakness. During the behaviorists' reign, evidence came to light that animals do indeed have mental representations that affect how they learn. The work of Edward Chace Tolman (1886–1959), a behaviorist himself, challenged some of the assumptions of traditional behaviorist doctrine. Tolman allowed rats to explore a **maze** and, in some cases, transported them through the maze. In both scenarios, the rats were subsequently able to navigate the maze successfully, even though they had never been rewarded for doing so. Tolman suggested that the rats developed a **cognitive map**, or a mental representation of the maze, that enabled them to navigate the maze without conditioning. This shows that animals can acquire behaviors through experience, without being reinforced—a concept called **latent learning**.

Tolman believed in five initiating causes of behavior: (1) the environmental stimuli (S), (2) physiological drive (P), (3) heredity (H), (4) previous training (T), and (5) age (A). Behavior (B) can then be expressed as a function of these in a simple equation: $B = f_x(S, P, H, T, A)$. But between these initiating causes and the final behaviors exist what he called

intervening variables. It is the intervening variables that are the determinants of behavior. They are the internal processes that connect the prior stimulus situation with the response. An example of an intervening variable is hunger. Although hunger cannot easily be observed, it can be related to other experimental variables that are readily observable, such as the amount of time that has elapsed since an animal last ate or how much food the animal consumes at the end of a period of not eating. Intervening variables were the behaviorists' attempt to objectify internal mental states. However, these states were acknowledged to exist only as they pertained to some external, easily measured variable, and in many cases, it was difficult to objectify them. Tolman himself later came to accept that a complete understanding of intervening variables was perhaps not possible.

OVERALL EVALUATION OF THE PSYCHOLOGICAL APPROACH

We have covered a plethora of theoretical positions in this chapter. There are several reasons for this. Psychology was the first discipline to systematically apply experimentation to the study of mind. It was thus a new discipline with many followers advocating many positions. Psychologists additionally had a very difficult task in front of them, which was to try to understand things that at the time could not be easily seen or measured. This lack of precision and early reliance on nonscientific methods such as introspection may have led to an overreliance on theory. Also, as was not the case in other disciplines, there was no overarching theory or framework for psychologists to work within. It would not be until the rise of the cognitive approach and the adoption of an information-processing perspective that some kind of unity would come to the field. We turn our attention to this cognitive approach in the next chapter.

SUMMING UP: A REVIEW OF CHAPTER 3

1. Psychology is the scientific study of mind and behavior. Throughout its relatively short history, this discipline has been characterized by the development of many competing theories of what the mind is.

2. In the scientific method, a hypothesis or conjecture about the world is tested using an experiment. In an experiment, an independent variable is manipulated and its effect on a dependent variable is measured. The results either support or fail to support the hypothesis. Over time, this process, when repeated, produces a theory—a more general understanding of how the world works.

3. The voluntarist movement marks the first time experimental methods were applied to mental phenomenon. Its progenitor, Wilhelm Wundt, also used introspection, or inward looking, as a means to catalog the elements of the mind.

4. The structuralists, led by Edward Titchener, had a research agenda similar to that of the voluntarists. They also attempted to catalog

mental elements and believed that they interacted with one another in much the same way as chemical elements do.

5. William James is considered the founder of the functionalist movement in psychology. In this view, the mind is best understood as a flow or ongoing process.

6. Gestalt psychologists, such as Max Wertheimer, wanted to know how mental parts combined to form wholes. They postulated the laws of perceptual grouping, which describe how we see visual objects.

7. Sigmund Freud started psychoanalytic psychology. He believed that the mind as a whole consisted of multiple different and competing minds that were in constant conflict with one another.

8. Skinner was one of the most prominent members of the behaviorist movement. According to this perspective, an organism's behavior is understood by looking at the environmental stimuli that come before it or following it. Punishments and rewards are thought to play a prominent role in shaping behavior.

SUGGESTED READINGS

Freud, S. (1962). *The ego and the id.* New York, NY: W. W. Norton.

James, W. (1890). *The principles of psychology.* New York, NY: Dover.

Koffka, K. (1935). *Principles of Gestalt psychology.* New York, NY: Harcourt.

Kohler, W. (1927). *The mentality of apes.* Berlin, Germany: Royal Academy of Sciences.

Myers, D. G. (2001). *Psychology.* New York, NY: Worth.

Schultz, D. P., & Schultz, S. E. (1987). *A history of modern psychology* (4th ed.). Orlando, FL: Harcourt Brace Jovanovich.

Skinner, B. F. (1971). *Beyond freedom and dignity.* New York, NY: Alfred A. Knopf.

Watson, J. B. (1970). *Behaviorism.* New York, NY: W. W. Norton.

CHAPTER FOUR

THE COGNITIVE APPROACH I

Vision, Pattern Recognition, and Attention

SOME HISTORY FIRST: THE RISE OF COGNITIVE PSYCHOLOGY

To understand the cognitive perspective, we need to backtrack to our prior discussion of behaviorism. You may recall from the previous chapter that for many decades, behaviorism was the dominant movement in psychology. Behaviorists avoided the study of mental phenomena because they believed that they were too difficult to define or measure. They stuck instead to external, observable behaviors, which were more amenable to scientific scrutiny.

Cognitive psychology can be seen as a backlash or counterrevolution to the behaviorist movement. This becomes evident when one examines the basic assumptions of the cognitive movement (Ashcraft, 2002). Whereas the behaviorists avoided studying the mental world, the cognitivists firmly acknowledged the existence of mental processes and focused their investigatory attentions on them. Whereas behaviorism saw the mind as a passive organ that operated according to the simple rules of conditioning, cognitive psychology saw the mind as active—selecting information from the environment, relating it to prior knowledge, and acting on the results of such processing.

In addition to this backlash against behaviorism, perhaps the most significant reason for the rise of cognitive psychology was the development of new technology. During the behaviorist era, there were no ways of "peering inside the head." That is, one could not directly measure mental processes. After the start of the cognitive movement, new technologies emerged that provided a more accurate picture of mental processes as they were occurring. The new devices included positron emission tomography, computerized axial tomography, and functional magnetic resonance imaging. These techniques are described further in the neuroscience chapter.

Learning Objectives

After reading this chapter, you will be able to:

1. Summarize how the subfield of cognitive psychology came about.
2. Critique Fodor's modularity of mind.
3. Compare template matching theory with feature detection theory.
4. Describe the three stages of David Marr's computational approach to vision.
5. Assess parallel and serial processing in Treisman's feature integration theory.
6. Contrast capacity-based models and selection-based models of attention.

Another important technological innovation was the development of the personal computer. The introduction of the transistor brought down the size and the cost of computers, making them available to more people, both inside and outside academia. This spurred psychologists to begin thinking more about them. Psychologists realized that the mind, like a computer, could be viewed as a device that represented and transformed information. The mind-as-computer metaphor was born. Computers thus accelerated the adoption of the information-processing view, not only in psychology but also, more generally, in other cognitive science fields.

THE COGNITIVE APPROACH: MIND AS AN INFORMATION PROCESSOR

So now that we know how cognitive psychology came about, what exactly is it? **Cognitive psychology** is the study of knowledge representation and use in human beings. It is concerned with understanding how people represent, process, and store information. According to Ulric Neisser (1967), one of the early founders of the movement, *cognitive psychology* refers to all processes whereby "the sensory input is transformed, reduced, elaborated, stored, recovered, and used" (p. 4). The many verbs used in the preceding sentence give us a sense of the many possible information-processing activities of the human mind.

Cognitive psychology differs from other approaches in cognitive science in that its focus is on human information processing (as opposed to animal or machine modes of information processing). Like the practitioners of many other disciplines in psychology, cognitive psychologists adopt the scientific method as their primary tool of investigation. Researchers test hypotheses by analyzing data from controlled experiments. However, cognitive psychology also supplements experimentation with modeling and computer simulation. A specific information-processing model of a mental capacity can be run on a computer. The results can then be compared with data from human experiments. This is often a synergistic and iterative procedure. Parameters of the computer model can be refined so as to provide a better fit between the computer model and the empirical data. Aspects of the simulation can also yield insights that cause researchers to go back and design new experiments.

A defining characteristic of the cognitive approach is the way it represents human information processing. These processes are often conceptualized using a **process model**. In a diagram of a process model, boxes are used to designate each stage or step in an information-processing sequence. Arrows that point toward or away from the boxes represent the flow of information among the stages. Many of the figures in this chapter show a process model that depicts a particular theory of human computation. Feel free to examine a few of them now.

Process models, in the classical view of information processing, carry two assumptions. First, they are assumed to be sequential, meaning that information that lies within one stage is processed before it is output to the next. Information cannot be processed simultaneously in multiple stages. Second, excluding inputs and outputs, processing that occurs within one stage is independent of processing that occurs within other stages.

These assumptions were later challenged by the connectionist view of information processing, which adopts a radically different architecture as the basis of cognition.

Process models are a very important part of the cognitive perspective. They are a powerful conceptual tool for understanding human information processing. In fact, the remainder of this chapter and the next are devoted to describing the major processing models that underlie various domains in cognition. These domains include perception, attention, memory, categorization, imagery, and problem solving. Humans can perform a wide range of cognitive processes, so many in fact that it is not possible to describe all of them in this book. In this chapter, we focus on visual pattern recognition and attention as these topics have been well investigated. We neither discuss other visual processes like color perception nor describe auditory perception. Similarly, in Chapter 5, we do not discuss reading or story comprehension. These are all very important, but entire courses can be devoted to each. In this and the next chapter, our goal is more limited: to demonstrate how major cognitive processes take place and how cognitive process models attempt to account for them.

MODULARITY OF MIND

An assumption of many cognitive theories is **modularity of mind**. According to this idea, the mind is made up of innate, functionally independent modules. The boxes in the process models that are described throughout this chapter and in the next chapter can in many cases be viewed as modules that carry out their cognitive processing independently of one another, not interacting in any significant manner. Because modularity of mind was such an important and pervasive assumption in the cognitive perspective, let's take some time now to examine it further.

Jerry Fodor (1983) was the most outspoken proponent of the modular view. In his book *The Modularity of Mind*, he gave a general description of the role these putative structures play. In his view, information arriving from the outside environment passes first through a set of sensory transducers. These convert the information to a code that is suitable for processing by the modules, which are domain specific—that is, able to handle information of a specific type only. The modules then convert the results of their operations into a common code that can be interpreted by other nonmodular, domain-general processors.

In the same volume, Fodor (1983) lists the many characteristics of a module. Modules are hardwired, meaning that they cannot be constructed from more basic units. They are genetically determined, domain specific, fast, automatic, stimulus driven, and not subject to control by a central authority. Modules are mandatory: They are triggered into operation by the presence of the appropriate information. Fodor also states that modules are informationally encapsulated. By this, he means that other mental processes can only have access to a module's output; they cannot influence or access its inner workings.

Evaluating the Modular Approach

Early case studies of brain-damaged patients appear to support the modular view. There are many case studies of patients who have suffered damage to a particular brain area

as a result of a stroke or an accident. These patients then suffer seemingly very specific deficits. In an aphasic patient, damage to one part of the brain might hamper the patient's understanding or comprehending language, whereas damage to another region might cause difficulties only in his or her speaking or in producing language. See the chapter on linguistics for more on these disorders. Similarly, there are patients with developmental genetic disorders in whom one language ability has been preserved while another has not. Individuals with Williams syndrome suffer grave deficits in visuospatial cognition but are relatively unaffected with respect to the processing of face-specific information (Bellugi, Wang, & Jernigan, 1994).

Skeptics counter these claims. They point out that in many brain-damaged patients, there is no clear-cut dissociation of one function from another (Tyler, 1992; Van Orden & Kloos, 2003). Also, genetic disorders of the sort described above rarely evidence highly specialized deficits—usually, more general impairments are manifest (Bishop, 1997). Evidence from studies of early brain development additionally fails to support claims of modularity. Experiments that have used brain-scanning techniques show that some language abilities are bilateralized—that is, located in both cortical hemispheres—early in development. Hemispheric specialization occurs later, which strongly suggests that these abilities cannot be genetically preprogrammed (Mills, Coffey-Corina, & Neville, 1993).

Current thinking is that the brain may contain a number of processing structures that are somewhat modular but that these may be a result of normal developmental processes. However, the extreme opposite view from modularity—that the brain is entirely domain general—does not appear to be tenable either. The task for future researchers in this area is to define which developmental factors shape the formation of modules and which cognitive functions are more general purpose and nonmodular in nature. The role that development and genes play in the creation of modules is discussed further in Chapter 8.

THEORIES OF VISION AND PATTERN RECOGNITION: HOW DO WE RECOGNIZE OBJECTS?

Perception is the process by which we gather information from the outside world via the senses and interpret that information. Most work in perception has focused first on vision and second on audition. Far less research attention has been given to the remaining senses. This bias is species centric, as vision is our most sophisticated sense and the one we rely on most. Many of the major perceptual theories are, therefore, based on and framed exclusively in terms of visual processing. In this section, we will examine several broad theories of human vision, as we acknowledge that much theoretical work remains to be done for the other senses.

One of the main functions of a visual system is the recognition of patterns. To demonstrate how important this is, imagine looking out at the world and not being able to recognize what you see. You would be able to navigate successfully—for example, you would be able to step out of the way of an oncoming car, but you would not recognize what just went by. You would be able to see your best friend but not know who he or she

is. You would be able to see written words but not read them. **Pattern recognition** is the ability to identify objects in the environment. We seem to do it effortlessly. Yet as we will see, this process is actually quite complex and far from being entirely understood.

Template Matching Theory

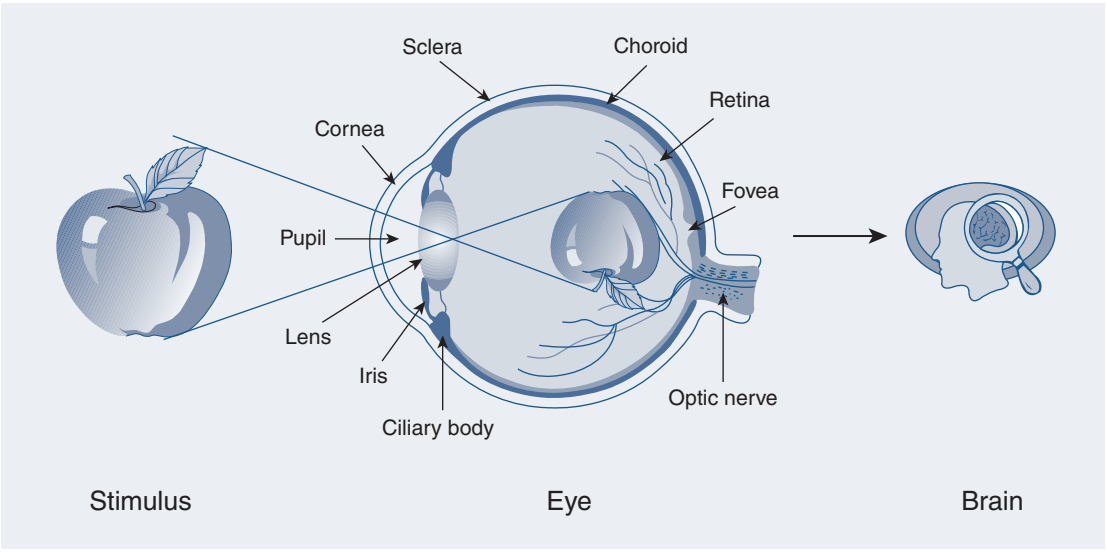
Any description of a pattern recognition process must begin with the stimulus. The stimulus is the actual object in the external world that we are trying to recognize. Light striking the stimulus is reflected and projects an inverted image of the stimulus onto the retina. The retina is a layer of photoreceptor cells that lines the inside of the back portion of the eye. The retina performs some preliminary processing of the image, after which information about the stimulus is passed posteriorly along a set of pathways toward visual areas in the brain. It is in these areas where the bulk of pattern recognition and other perceptual processing takes place. Figure 4.1 shows an overview of these steps.

According to the template matching theory of pattern recognition, an image generated by an external stimulus is matched to an internal mental representation of the stimulus, called a **template**. The degree of overlap between the image and the template is then computed. This overlap is a measure of how similar the two are to each other. A high degree of overlap will produce recognition of the object. Note that both the perceptual image formed from the world object and the template drawn from memory are examples of visual representation. They represent the object but in different ways. The image is constructed from visual perceptual processes and is more of a precise visual replica of what is "out there," its purpose being to reproduce as carefully as possible its visual appearance. The template is an internally constructed representation drawn from information in long-term memory and is based on past experience. In the model, they differ from verbal symbolic representations of the sort found in linguistic processes. The computational processes in this model are the construction of the image and the template and the matching between them. Template construction would require search and retrieval of information from long-term memory storage. Matching would require some sort of "difference operator," a way of comparing the two to determine their differences.

Evaluating Template Matching Theory

The problem with this approach is that there is a wide range of variation and possibility for any individual stimulus. Take the simple case of the letter *A*. It can vary in terms of its location in one's visual field, its size, shape, and orientation. For template matching to work, there must be a separate template for each of these possibilities. Imagine how many templates would be required for just one letter of the alphabet. We would need a template for an *A* that appears to the left or the right side of the visual field, other templates for big and small *A*s, others for *A*s written in script or appearing in different fonts, and yet others for *A*s that are placed horizontally or upside down. The number of possible templates for an individual object multiplied by the number of possible objects one could recognize becomes so large that it would become impractical to store templates or to use them effectively in memory. Figure 4.2 shows the template matching process and some examples of the ways in which a stimulus can vary.

Figure 4.1 Early steps in the recognition of a stimulus object.



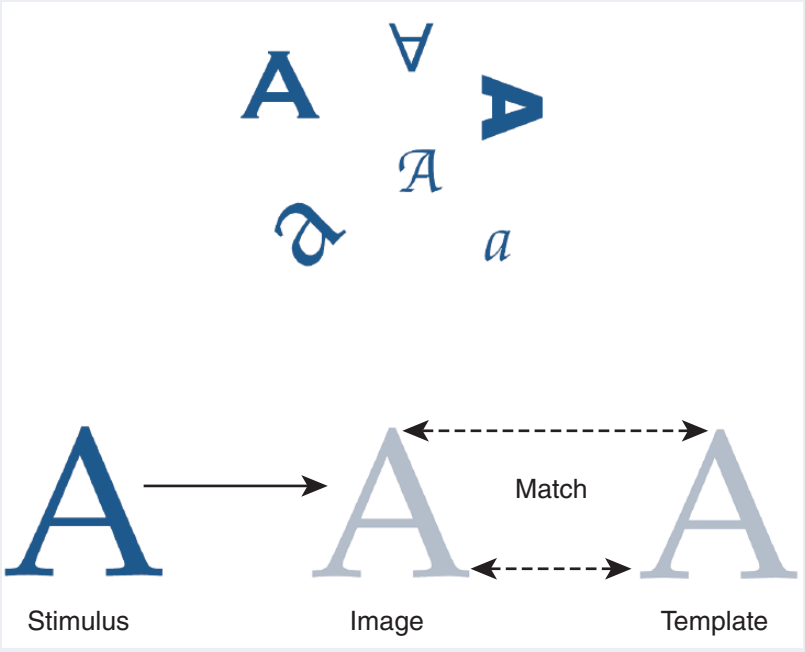
For this reason, template matching was relatively quickly dismissed as a theory of the way in which humans recognize patterns. It is considered a "strawman" theory, in that it is insubstantial. As a theory, though, it helps us conceptualize what is required of, and what gets in the way of, a pattern recognition process. Several computer versions of template matching have been successfully implemented. One is designed to read checking account numbers off the bottoms of checks. In these cases, the procedure works because variability in the stimulus has been reduced. The sizes, styles, and numbers of the digits that must be matched are relatively unvarying. These template matching systems would fail under conditions of handwritten digits, in which stimulus variability is much greater.

Feature Detection Theory

With feature detection models, an image of a stimulus, rather than being matched in its entirety to a template, is broken down into its component features. A feature is a part or a subset of an object. The idea is that each different combination of features uniquely specifies a different object. The upright letter *A* can be specified by the fact that it has a short horizontal line as one of its features and two longer diagonal lines as additional features. It can be distinguished from the upright letter *B*, which has a long vertical line and two short loops as features.

The best-known feature detection model is called the **pandemonium model** (Lindsay & Norman, 1972; Selfridge, 1959). It gets its name from little mental "demons"

Figure 4.2 The top panel provides various depictions of the letter *A*, all recognizable. The bottom panel shows the sequence of events in the template matching model.



that represent processing units. These demons "shout" during the recognition process (creating a pandemonium). Figure 4.3 depicts a representation of the model in terms of how it might be used to identify the letter *R*. First, the stimulus, a letter *R*, is presented. It is represented by an *image demon*, which preserves the overall appearance of the letter. In the next step, there are *feature demons*. There is one feature demon for each possible stimulus feature. A feature demon shouts if he sees his own feature in the image. The upright letter *R*, in this example, can be thought of as having a vertical line, a diagonal line, and a loop and so has three feature demons. These feature demons would shout out in the presence of their own features. In the next step, there are *cognitive demons*, one for each possible letter. If they hear any of their corresponding features, they too will shout. The cognitive demon with the most features shouts the loudest. Finally, there is a *decision demon*. It listens to the cognitive demons. Whichever one shouts the loudest is chosen by the cognitive demon as the recognized letter.

Each of the demons in this model is a representation. However, given the advances in neuroscience, we have a better understanding of what they might be like. It is now known that simple features of the sort discussed here, like lines and angles, are coded

for by neurons in area V1 of the visual cortex and that more complex features made up of their combinations are also represented neurally higher up in the visual processing stream. Neurons that respond to letters or other complex images like faces can be found in area IT of the temporal lobe. The more complex features are constructed through a process of convergence, whereby lower-level features that make up their component parts send connections forward. This process is described in the network chapter and involves hierarchical organization. A summary of the process is outlined in Figure 7.12.

Evaluating Feature Detection Theory

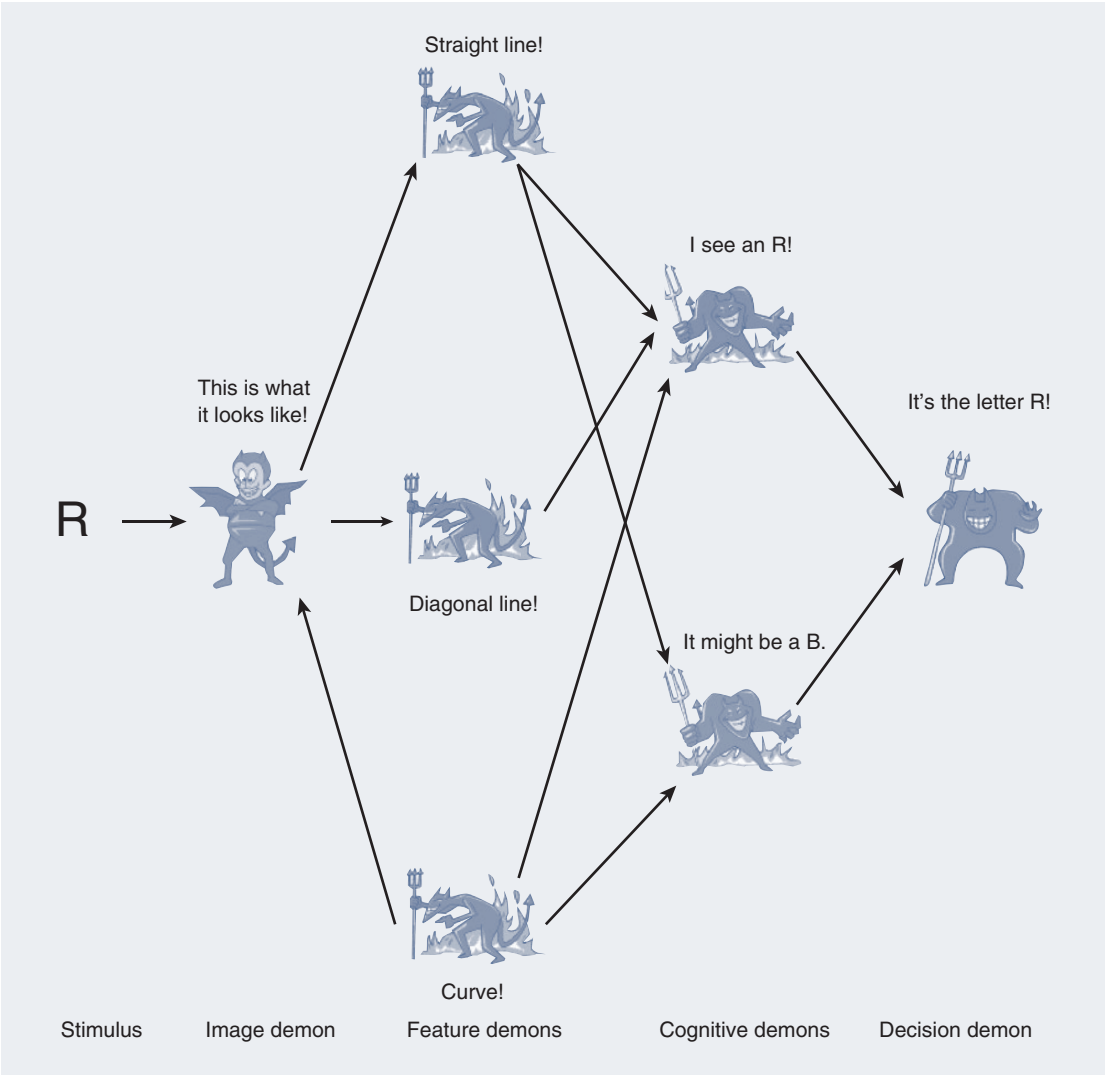
The pandemonium model of pattern recognition represents a significant improvement over template matching. It doesn't require an extensive number of templates and will have only as many feature demons as there are features and as many cognitive demons as there are letters or other objects. The model can also explain the types of mistakes that people make during recognition. Often, an individual will confuse two visually presented letters that have features in common. Continuing with our example, a person might mistake the letter *R* for the letter *B* because both have a vertical line and a loop. The decision demon might accidentally choose the cognitive demon that represents the *B* because it will be shouting almost as loud as the *R* cognitive demon.

Another reason to like the pandemonium model comes from neurophysiology. Evidence from this field demonstrates that neurons in the visual system act as feature detectors. Single-cell recordings of neurons in the primary visual cortex of cats show that these cells respond selectively to different features, such as a line of a given length and orientation (Hubel & Wiesel, 1962).

But feature detection models are not without their share of difficulties. For one thing, there is no good definition of what a feature is. For the letter *R,* the vertical and diagonal lines could together form the single feature of an angle, as opposed to their being separate features. Feature detection models are also bottom-up or **data-driven processes**. This means that they are driven entirely by the physical characteristics of the stimulus itself and fail to take into account the larger context or meaning. In real-world recognition scenarios, objects appear in varied contexts. The surrounding conditions often yield clues that can aid identification. In Figure 4.4, the central figure can be perceived as two different symbols, depending on the context to which one pays attention. When viewed as part of the vertical arrangement of numbers, it is perceived as the number 13. However, when viewed as part of the horizontal arrangement of letters, it may be perceived as the letter *B.* This process is called **pattern completion**, where the mind fills in the gaps of an ambiguous or incomplete stimulus based on the surrounding context.

Additional evidence for the role of context in recognition comes from what is called the word superiority effect (Reicher, 1969). Here, the time it takes to recognize the letter *A* in CAT is shorter than the time it takes to recognize it alone or as part of a nonsense word, such as YAX. This suggests that the overall word is processed before any of the letters. Because the letter *A* appears in the word CAT, the presence of the entire word facilitates recognition of the individual letter. Both the example of the ambiguous figure given above and the word superiority effect are examples of top-down or **conceptually driven processes**, in which context and higher-level knowledge aid recognition.

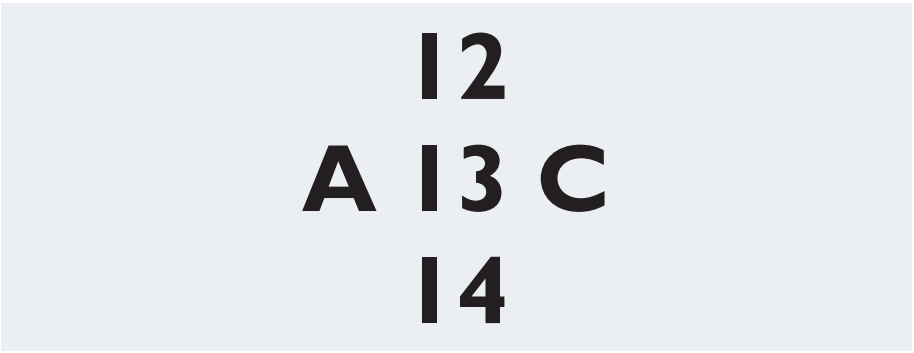
Figure 4.3 The pandemonium model of pattern recognition. Different "demons" perform different steps in the recognition of the letter *R*.



Recognition by Components Theory

Stimuli as they are represented on the retina are rarely viewed in the same way again. As we discussed in the template matching model, an object may appear in different locations, sizes, orientations, and shapes, to name just a few possible types of variation. Any pattern recognition mechanism must, therefore, produce a description of the object

Figure 4.4 Effects of context in perception. What is the central figure?



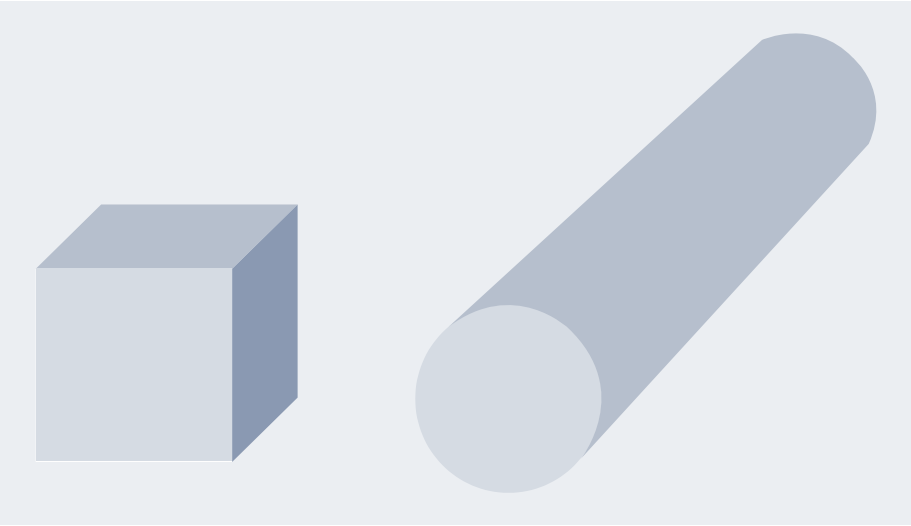
that is impervious to these kinds of changes. This is known as the **shape constancy** problem in perception. Somehow, our visual system must extract from the stimulus input aspects that are invariant with respect to viewing angle, distance, lighting, and so on. The template matching solution was to store multiple copies of the object. The feature detection solution was to extract certain invariant features—aspects of the object that remain constant.

The psychologist Irving Biederman (1995) also proposes a feature extraction account for the constancy problem; his approach is interesting because it specifies how shape constancy may be achieved for three-dimensional (3-D) object representations. In his theory, features gathered from a stimulus are recombined to form an object representation. These features he calls **geons**. A geon has a basic volumetric shape—for example, a cube or a cylinder. Altogether, there are 24 types of geons in his theory. Figure 4.5 shows two examples of individual geons.

Geons are said to have three basic properties. The first is view invariance. This means that geons can be identified when viewed from many different perspectives. The second is discriminability: One can tell one geon apart from another no matter what the viewing angle. The fact that geons have these characteristics supports the notion that they may be used as complex features in human pattern recognition since they can account for the view-invariant aspect of shape constancy.

The third property of geons is resistance to visual "noise." Geons can be perceived even when many of the contours that make them up are obscured. In one study, Biederman (1987) showed observers line drawings of objects that were masked by noise. The noise in this case was amorphous black regions resembling patches of spilled ink. In one condition, the noise covered the object contours in a way that preserved the geons. The noise for the most part in this condition covered single stretches of contour. In a second condition, the noise covered the contours such that the perception of the geons was disrupted. The noise here covered contour junctions, which are needed for the construction of the surface planes of 3-D shapes. Observers had no trouble recognizing the shapes in

Figure 4.5 Cubes and cylinders can be considered examples of geons. These can then be assembled to form more complex objects.



the first condition but had great difficulty doing so in the second. These results suggest that geons may help us recognize objects that have undergone occlusion, a partial covering up by other objects, since in the real world, occluding shapes rarely obscure junctions.

Geons are examples of very high-level representations. The visual input presented to the retina is two-dimensional (2-D), but as we are all aware, what we perceive is 3-D. The visual system uses cues in the environment to determine what objects and parts of objects are closer or farther. For instance, objects or parts that are larger are generally closer. It also uses the disparity, or difference of view between the two eyes, to tell us what is close or far. This sort of processing, known as **depth perception**, is a necessary part of geon formation. The advantage of a 3-D representation is obvious; it is a more flexible mental depiction of the object and allows us to manipulate and interact with it more successfully. The way in which individual geons may be assembled to form more complex objects is outlined below in David Marr's (1982) computation model of pattern recognition.

Evaluating Recognition by Components Theory

Biederman's recognition by components theory is supported because it can account for view invariance and occlusion. Another strength of the theory is the small basic feature set that consists of just 24 geons. In other pattern recognition theories, features are either unspecified or extremely numerous. Critics of the theory point out that broad classes of objects can be discriminated on the basis of geons but that fine-level object categories cannot. For example, geons can explain why we have no problem telling the difference

between an airplane and an automobile, but they have difficulty in providing an explanation for how we can also tell apart objects that are much more similarly shaped, such as two similar bird species (Perrett & Oram, 1993). For these sorts of differences, we must rely on more fine-grain continuous-valued features, such as the angle of an object part or perhaps color and texture (Edelman, 1999).

INTERDISCIPLINARY CROSSROADS: COMPUTATIONAL VISION AND PATTERN RECOGNITION

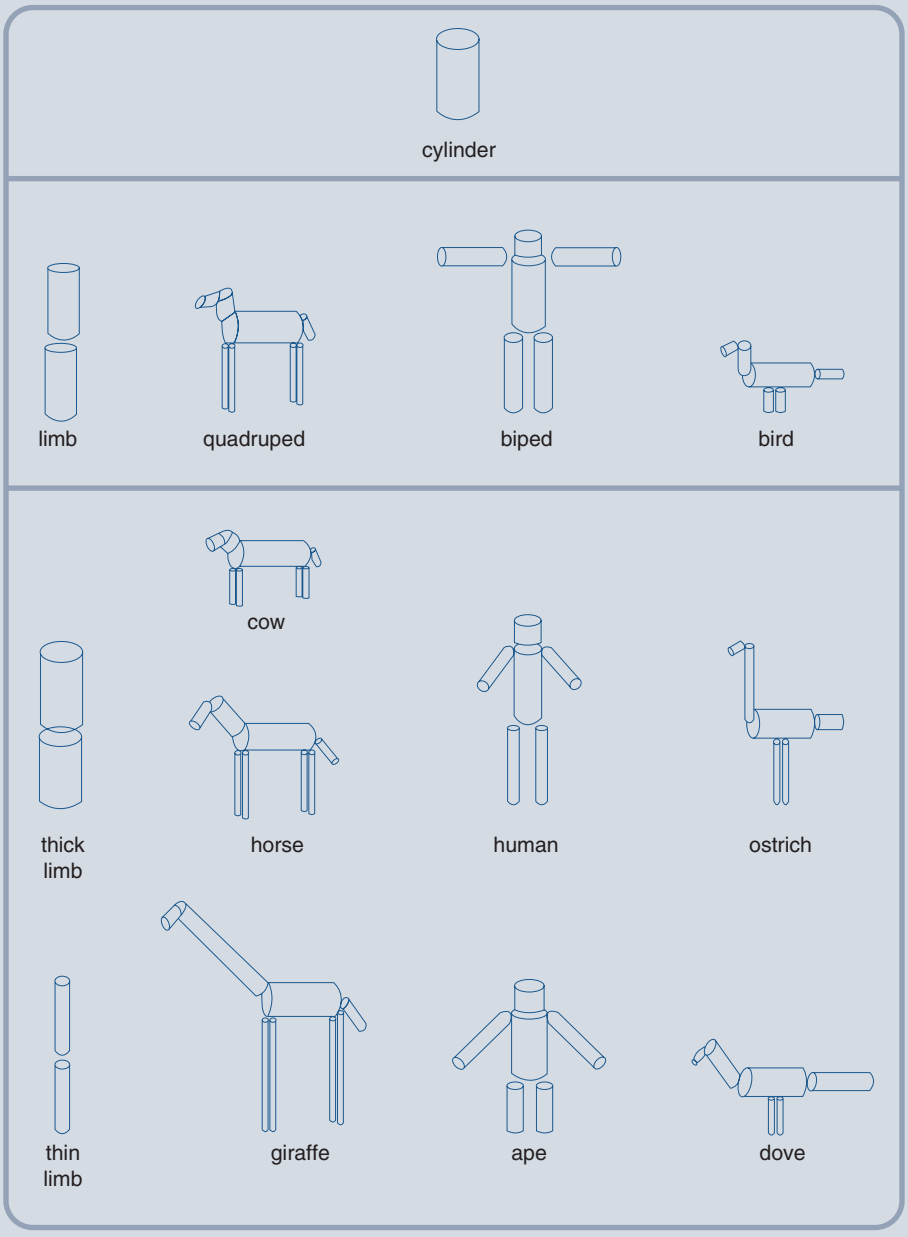
In 1982, the renowned cognitive scientist David Marr wrote a book that described his influential theory of pattern recognition. He adopted a computational approach to vision. In this approach, the visual system is treated as a computer. The "computer" receives an image of an object and, after performing several algorithmic processing steps, accomplishes recognition of the object. Each major processing step produces a more fully articulated version of the object—what Marr calls "sketches."

In the first stage, the image of an object is presented to the retina. This image is analyzed in terms of the intensity values, or areas of light and dark, that make up the image. Adjacent regions of sharp contrast in the image (that go from light to dark or vice versa) indicate the presence of edges and contours. The edges and contours in turn determine the basic features of the object's shape. In his scheme, these features include line segments, circular shapes, and patches of color. The result is a **raw primal sketch** of the image. To get something of an idea of what the raw primal sketch is made of, imagine standing too close to an impressionist painting by Monet or Van Gogh. You'd be able to see the patches of color and small portions of contours, but the actual shapes and objects would be hard to see.

The raw primal sketch then undergoes further processing. Features that are similar in size and orientation get grouped, in much the same way that parts are grouped into wholes according to the Gestalt principles discussed earlier. The groups of features are then processed again to produce a representation of the object that includes its surfaces and approximate layout. Marr calls this the **2½-D sketch**, because it is more than the 2-D primal sketch, but it is not quite a full 3-D perception of the objects. To get something of an idea of what the 2½-D sketch is made of, imagine standing too close to a Picasso portrait. You'd be able to recognize an eye, another eye, a nose, a mouth, but the normal 3-D layout of a recognizable face might be somewhat hard to perceive.

This image is next transformed into a complete 3-D representation. In this **3-D sketch**, the axes of symmetry and elongation link the resultant object parts (see Figure 4.6). A symmetry axis is a line that divides an object into mirror-image halves. An elongation axis is the line defining the direction along which the main bulk or mass of a shape is distributed. These axes serve to hierarchically organize the parts into a coherent, large-scale object that can be recognized. According to Marr, it is this final 3-D construct that we perceive and are consciously aware of.

Figure 4.6 Object representations in Marr's 3-D sketch.



Source: Marr & Nishihara (1978). Reprinted with permission.

In Marr's model, the 2½-D sketch contains a **viewer-centered description**. This is one that is particular to the viewer's point of view. Because the viewer's location, perspective, orientation, and so on are subject to change, object recognition under these conditions is difficult. But Marr's 3-D sketch has an **object-centered description**. The object's parts are described relative to one another and are linked on the basis of shared properties and axes. The relationships between parts thus remain intact across object transformations due to changes in viewing angle, thus solving the shape constancy problem and enabling recognition under a broad variety of conditions.

Whereas the template matching model focuses on the form of representations in pattern recognition, Marr's model focuses on the computational steps or processes involved. Notice here that we get a detailed step-by-step description of what would transpire for an object to be recognized. Each step is a computation—from the determination of intensity values, to edge extraction, to basic feature determination, to the grouping of features, to a sketch of surfaces and the overall layout, to the formation and linkage of 3-D object parts. This is an algorithmic approach, one that can be implemented both in a neural system like the brain or a computer software program.

Evaluating Marr's Computational Approach to Vision

It is important to point out that Marr's theory is not a statement of how the visual system actually recognizes objects but how it *might* go about doing so. Although it is based on existing experimental and neurophysiological evidence, Marr's theory is more a speculation on the processes required of any pattern recognition mechanism, human or machine. We should also emphasize that Marr's theory involves feature detection, as the image is first broken down into component features that are then reassembled to form a complete, holistic object.

Feature Integration Theory

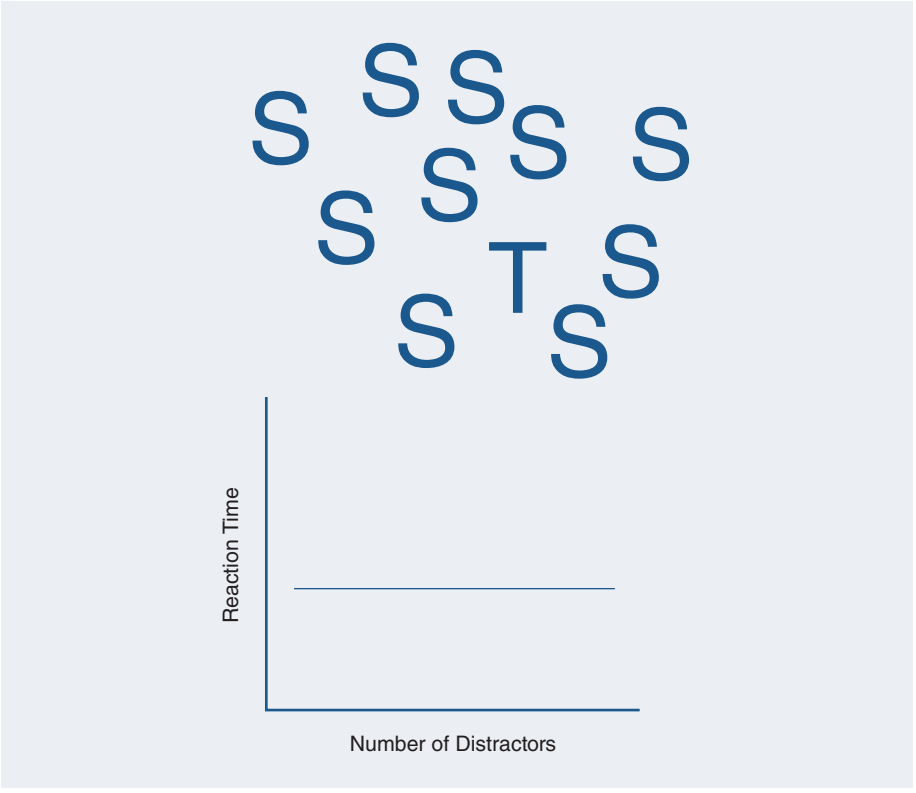
Another theory that entails feature extraction and recombination goes by the name of feature integration (Treisman & Gelade, 1980). This theory is unique because it brings to light the special role that attention plays in pattern recognition. It is based on a rather extensive set of cognitive psychology experiments, some of which we will describe below. Feature integration, like Marr's computational view, is also a stage-based theory. An object's basic features are identified by way of the image in the **preattentive stage**. These features can include color, motion, orientation, and curvature. Then, the features are combined during a **focused attention stage**. Following this, the object is recognized.

Let us examine more closely how features are identified and combined in these two stages. In a series of clever experiments, Treisman asked participants to report if a target item was present in a visual field that was filled with nontarget items, or distractors (Treisman & Gelade, 1980). This task has since become known as **visual search**. On half of the trials, there is a target item present, and on the other half, the target is absent. Importantly, the number of distractors can vary. For example, in one trial, the target might be the single letter *T*, hidden among five *S*s. In another trial, the *T* might be hidden among 10 *S*s. In each trial, the time it takes for the participant to locate the

target is recorded. A plot that shows the time needed to identify the target as a function of the number of distractors is then made. Figure 4.7 shows a display and the plot of the search function.

For the sort of task described above—locating a *T* among *S*s—the search function is flat. The amount of time it takes to locate the target is small and constant, regardless of the number of distractors. Having 10 distractors does not make locating the target more difficult than having only 5. Participants in the experiment report that, under these circumstances, the target seems to "pop out" at them, so the phenomenon has since been identified as perceptual **pop out**. Treisman argues that pop out is preattentive. A preattentive process is one that happens automatically and effortlessly, before attention would have to be applied. In vision, preattentive processes that are part of the processing of an image happen very quickly—usually within the first 100 milliseconds. The participants

Figure 4.7 **Location of the target letter in the top half is an instance of preattentive and parallel search. The target T seems to "pop out" effortlessly, even if more distractors are added. The bottom half shows an idealized version of the flat search function for this type of task.**

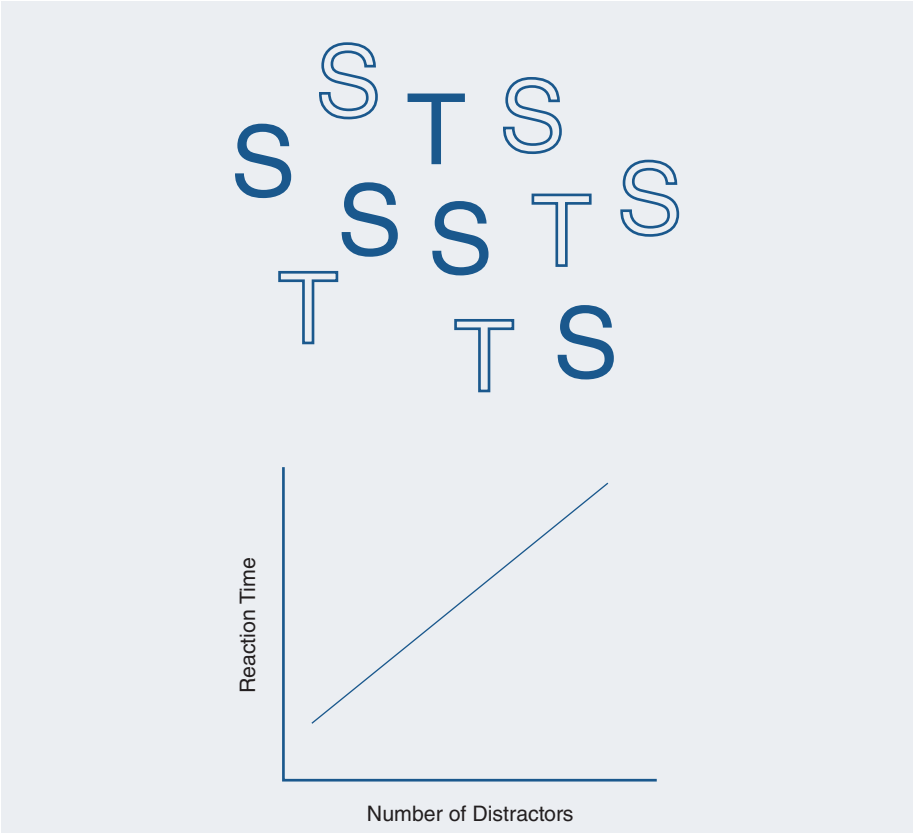


have no voluntary control over the process and don't need to proactively search through the display to locate the target. According to feature integration theory, an array of feature detectors is prepared to detect the single feature of curvature anywhere in the visual field simultaneously, or *in parallel*. The *T* is composed of straight lines with no curvature, whereas the *S*s are all curved, so the single visual feature of curvature is sufficient to distinguish the target from the distractors. Pop-out also happens when the target and the distractor differ in color, motion, or brightness, suggesting that these dimensions constitute a basic set of visual feature maps, much like those proposed in David Marr's raw primal sketch stage. Treisman calls this **parallel search** because the target seems to be scanned in parallel across all the display items at once. Parallel search is characterized by a flat search function. It occurs when the target differs from the distractors along a single perceptual feature. In this example, the relevant feature is curvature, but in the next example, there are two relevant features.

Now imagine a second type of visual search (see Figure 4.8). This time, your job is to locate a blue *T*. It is hidden among white *T*s, as well Ss and blue *S*s. Thus, curvature alone will not distinguish the target from the distractor, and neither will color alone. There are multiple noncurved items and multiple blue items. In this case, the feature of noncurved and the feature of blue must be conjoined in order to find the target. As a result, the search function shows a linear increase in reaction time as a function of the number of distractors. The greater the number of distractors, the more time it takes to identify the target. This suggests that a different type of search strategy may be involved. Participants in this instance report having to proactively look around the display until they find the target. In fact, the increase in reaction time generated by each additional distractor is such a steady fixed amount that the search function tends to be highly linear and straight. It is almost as if the cognitive mechanism performing the visual search is sequentially checking each display item one at a time to see if it has the conjunction of target features (e.g., noncurved and blue, in this case). For this reason, Treisman calls this **serial search**. Serial search is characterized by a linear increasing search function and appears to require focused, concentrated attention that is under voluntary control. The search process seems to involve matching each item with respect to two dimensions. If the item fails to satisfy either of these conditions, the process must then direct attention to the next item and do the same thing, over and over, until the target is located. Serial search is necessary when the target differs from the distractors along two perceptual dimensions, such as shape and color. Treisman says that the job of voluntary attention, in this situation, is to "glue" together the different features. This attentional glue binds features together into coherent objects—a necessary step before object recognition can take place.

So let's review what we know of Treisman's theory up to this point. In the preattentive stage, an object's features are identified. This process is fast, automatic, and effortless. In other words, it does not require attention and happens before attentional processes can even be brought to bear. A blue *T* in this stage is immediately broken down into its basic components of the color blue and its particular line segment orientations. In the subsequent focused attention stage, concentrated attention is required to glue these different features together in one place so that they form an object representation. The color and line segments are effectively linked together by attention in one location in

Figure 4.8 **Location of the target letter in the top panel is difficult in this example of attentive and serial search. Attention must be paid to each item in the display. The search function for this task, an idealized version of which is shown in the bottom panel, is linear and increasing.**



the visual field. Only the combination of such features can tell us what an object really is and whether it differs from other objects with which it shares some of those features.

You may encounter a real-life example of visual search as you're about to leave the shopping mall. You look out into the parking lot to try to find your car, a red Honda Civic. If your car is the only red car or the only Civic of the same model as yours in the lot, it will pop out, and you will locate it immediately. If, however, there are other red cars and/or other Civics of different colors, you will have to focus attention on each car in the lot before you will be able to locate yours.

Evaluating Feature Integration Theory

Feature integration theory was highly influential in the community of perceptual researchers. Investigators conducted numerous experiments that generated replications and extensions of its basic findings. However, a number of findings have called it into question. For example, certain conjunctions of features can actually produce a flat search function with the experience of "pop-out," which shouldn't happen if it is a serial process (Nakayama & Silverman, 1986). Also, partial similarity among features affects the slope of the search function, such that searching for a red square among pink squares can actually produce data that look like a serial search process, even though it is a single feature difference (Duncan & Humphreys, 1989). And search functions can have slopes that are flat, steep, or anywhere in between (Olds, Cowan, & Jolicoeur, 2000; Wolfe, 1998), suggesting that there may not be two separate cognitive mechanisms for preattentive parallel search and for attentive serial search. Thus, while Treisman's stage-based theory for how visual search works has helped move research forward, this computer metaphor may not be quite as powerful as a neural network account of visual attention (Desimone & Duncan, 1995; Reali, Spivey, Tyler, & Terranova, 2006).

THEORIES OF ATTENTION: HOW DO WE PAY ATTENTION?

We have already alluded to attention, having described the role it plays in the process of combining features for visual pattern recognition. In this section, we will examine the broad nature of attention and describe several theories of how it is used in a wider cognitive context.

Attention is defined as concentrated mental activity. In general, we can think of attention as a form of mental activity or energy that is distributed among alternative information sources. Informational sources can be stimuli from the outside world or thoughts or other forms of internal mental content. When attention is allocated to one source, that source comes into conscious awareness, and the processing of that information is usually facilitated. From a cognitive perspective, attention can be considered as having two sets of opposing characteristics. It is selective but divisible, as well as shiftable but sustainable. Let's discuss each of these.

By **selective attention**, we mean that attention can be focused onto one source of information and away from another. For example, in class you can decide to listen to the sound of the instructor's voice or to some competing sound, such as the conversation of two students in the row behind you. At the same time, attention can be split or **divided** among several alternative information sources. You could, with some effort, listen to the voices of both the instructor and the students. In this situation, the ability to process information from the two sources would be significantly reduced. Generally, the greater the number of sources among which attention is divided, the less attention there is to devote to any one of them.

Implicit in the notion of selectivity is **shiftability**. One can selectively attend to information source A while ignoring B, then switch back to B and ignore A. Or one could

attend source A, then B, then C, and so on. The point is that attention can be repeatedly shifted among different sources. This shifting can be voluntary or involuntary. Distraction is an example of the involuntary shifting of attention to some irrelevant or undesirable piece of information. The converse of shifting is **sustainability**. Here, attention is focused exclusively on one source and sustained in this way over time. We can think of sustained attention as extended focus or concentration.

Two general classes of theories attempt to explain attention. **Bottleneck theories** describe why it is that of all the information presented to us, only a small portion actually gets through. These theories attempt to explain the apparent bottleneck effect or the narrowing down of the information that reaches conscious awareness. Bottleneck theories are inherently theories of selective attention because they describe how some information is selected for processing as the rest gets discarded. Broadbent's filter model, Treisman's attenuation model, the Deutsch–Norman memory selection model, and the multimode model are all bottleneck theories. **Capacity theories**, on the other hand, are essentially theories of divided attention. They conceptualize attention as a limited resource that must be spread around different informational sources. Kahneman's (1973) capacity model is an example of a capacity theory.

Broadbent's Filter Model

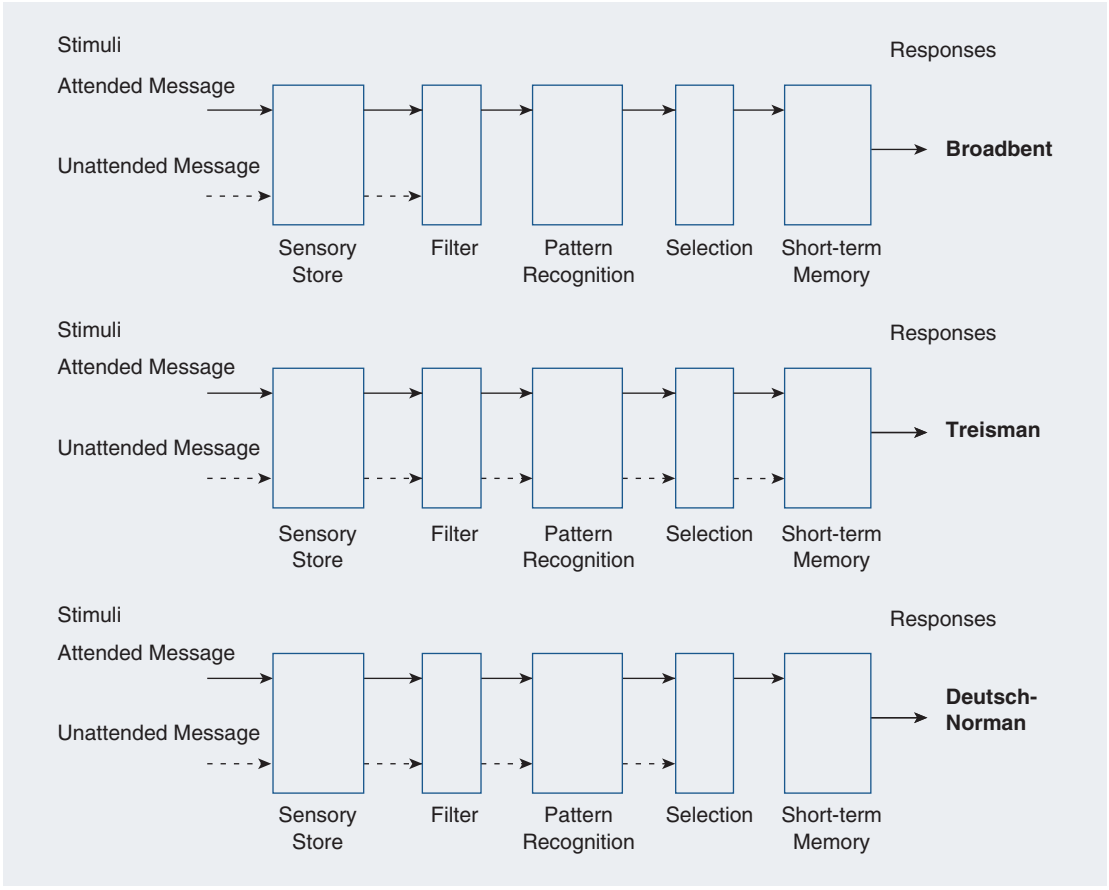
The British researcher Donald Broadbent did some of the earliest pioneering work in cognitive psychology. In his 1958 book, *Perception and Communication*, he proposed a theory to account for why we can't follow two streams of information coming in simultaneously via the two ears. Before we get to the details of his filter model, we need to describe some of the experimental evidence that led to it.

The **dichotic listening task** has been widely used to study selective attention. In this paradigm, a participant wearing headphones listens to two different messages being played, one over each ear. For instance, a male voice may be reciting one list of words into the left ear as a female voice recites a different list of words into the right ear. The participant's task is to pay attention to, or shadow, one of the voices and to ignore the other. The messages, or information sources, presented to the participant through both ears are in this context referred to as channels. Afterward, the participant is asked to recall what he or she has heard, for both the shadowed and unattended channels.

Early dichotic listening studies showed that participants were fairly accurate in recalling content from the shadowed ear but quite poor in recalling that coming from the unattended ear (Cherry, 1953). To Broadbent, this suggested the presence of a filter that could block out messages that were being relatively ignored, allowing only attended information to enter awareness. Figure 4.9 shows a diagram of Broadbent's filter model. Information from both channels first arrives in sensory memory. This is a place where incoming sensory information is briefly held before it is processed. Next, a filter is applied to this information. The filter selects certain portions of the information—that coming from the shadowed channel in this case—and allows it to pass through. Information coming from the unattended channel is blocked.

The selection in Broadbent's model is based on the physical characteristics of the stimulus, such as location (left or right ear), pitch, and loudness. This conclusion was

Figure 4.9 Information-processing models of selective attention. The solid arrows indicate information from the attended message. The dashed arrows indicate information from the unattended message.



Source: Adapted from Reed (2000).

based on the fact that one could shadow messages easily using these as criteria. In the next step, whatever information gets past the filter then undergoes pattern recognition. Following this step, information travels past any selection mechanism to a short-term memory store, where it is held for a longer period of time and made available for subsequent processing and response. Notice that selection in this model is performed entirely by the filter and not by any selection mechanism that comes later. We describe this selection mechanism only to show how this model compares with others. Broadbent's model is referred to as an **early selection model** because the filter screens out information before it can be recognized.

Evaluating the Filter Model

Almost from the get-go, it was realized that there were some problems with Broadbent's filter model of the theory. One problem is summed up in what is one of the better names for a psychological phenomenon, the so-called **cocktail party effect**. Imagine that you are attending a party and chatting with several people. Somebody across the room then mentions your name. To your surprise, you realize that you heard it, even though you were completely engrossed in the immediate conversation. Researchers have empirically validated this intuitive finding (Moray, 1959; Wood & Cowan, 1995). The observed cocktail party effect and other studies have demonstrated that some information gets through the unattended channel. The filter does not block words of great personal relevance, such as your name and words associated with danger—for example, *fire*. Broadbent's model has the filter completely blocking all information coming from the unattended channel. Clearly, his view requires some revision.

Treisman's Attenuation Model

Attenuation theory is similar to filter theory but with one exception (Treisman, 1964). Instead of completely blocking out the unattended message, the filter attenuates or weakens it. Message attenuation is like having water running through two faucets, where each of the faucets constitutes a different information channel. The faucet corresponding to the shadowed message would be open all the way, allowing water to pour through. The faucet corresponding to the unattended channel would be only partially open, allowing just a small amount of water through. The addition of Treisman's model to Figure 4.9 would have the unattended message passing through all the processing stages, but in a weakened form.

Treisman explains that different words have different chances of making it through the unattended channel, due to a threshold effect. A **threshold** is the minimum amount of activation required to produce conscious awareness of a stimulus. Stimuli with associated low thresholds easily make their way into awareness. Those with associated high thresholds do not. A word's meaning determines its threshold. Important words and those with personal relevance, such as your name, have a lower threshold for recognition and make it past the filter. Less important words—for example, *chair*—have higher thresholds and are filtered out.

The Deutsch–Norman Memory Selection Model

The cocktail party effect shows that meaning, or semantics, is another criterion according to which information is selected. This runs contrary to early filter models that posit physical characteristics such as loudness or pitch as the primary basis for selection. Deutsch and Deutsch (1963) and Norman (1968), therefore, proposed a second selection process, one that is based on semantic characteristics. This selection happens later in processing. Their model is an example of a **late selection model**. The first stages of processing in this model are the same as those of the other models. Information from the sensory store is filtered on the basis of physical characteristics and then recognized.

However, before being passed into short-term memory, it goes through a secondary selection mechanism. This mechanism selects information on the basis of semantic characteristics or message content. The selected items end up in short-term memory and awareness. Those items not selected never reach an awareness. The Deutsch–Norman model shown in Figure 4.9 depicts information from two channels reaching the selection mechanism, which makes the choice as to which information gets through.

The Multimode Model of Attention

So where does selection take place? More recent perspectives propose that selection can occur early or late, depending on the situation. Michael Posner and Charles Snyder (1975) advocated the idea of a "movable filter," one that can operate at various points in processing according to the observer's needs. Data from another study support what seemed to some to be the case all along, that selection can be based on physical properties of the stimulus or on semantic properties (Johnston & Heinz, 1978). They found, however, that semantic selection imposes a greater cost—that is, it requires greater attentional resources than physical selection. The model that arises from the view that selection can be based on multiple modes—for example, the physical and the semantic—is called the **multimode model of attention**.

There is also neurophysiological evidence to support the multimode model. Hillyard, Hink, Schwent, and Picton (1973) used an event-related potential (ERP) technique to investigate when selection occurs. The ERP records brain electrical activity that happens in response to the presentation of a stimulus. According to their design, the researchers had participants attend to a stimulus—say, coming from the left ear—and compared the activity observed under that condition with observable activity related to an identical unattended target—say, coming from the right ear. They observed a very rapid change in electrical activity only 80 milliseconds after the stimulus was presented, which suggested that selection was happening early on in response to perceptual characteristics. However, other ERP studies showed a change in electrical activity that happened approximately 300 milliseconds after stimulus onset, which suggested a late selection based on task relevance and, perhaps, semantic features (Luck & Hillyard, 1994). To summarize, it appears that attentional selection is more flexible than was first assumed. It can happen early or late and can be based on more than just a single mode.

All of the selective attention models discussed thus far are based on the same general information-processing mechanisms. They all involve the same stages of cognitive processing that occur in the same order. The sensory input is represented in a sensory store, then filtered, recognized, selected, and represented again in short-term memory. Since they are assemblies of individual processes that minimally interact, they are instances of a box-and-arrow cognitive architecture like others that show up throughout this textbook. Each processing stage by itself consists of a complex set of computational processes that are not elaborated on here. The stages can be considered to be modules as each takes in information from a previous stage, performs a particular computation, and then passes the output of that computation on to the next stage.

Kahneman's Capacity Model of Attention

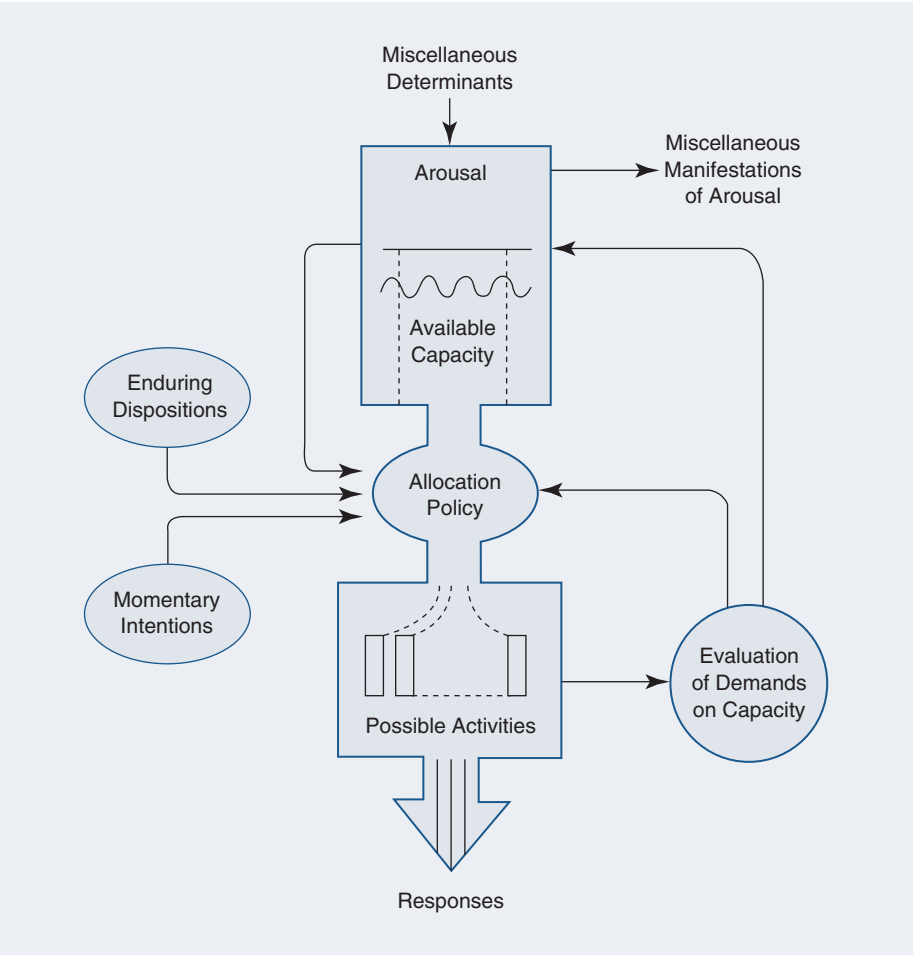
Capacity models describe attention as a resource. We can think of this resource as the amount of mental effort or energy required to perform a task. Like many other resources, attention exists in limited supply. At any given time, we have only so much of it that is available for use. Daniel Kahneman (1973), in his book *Attention and Effort*, outlined a detailed capacity model of attention. In it, he described the factors that affect available attentional capacity as well as the factors that determine how this capacity gets used. Because we are talking about the distribution of attention to different sources, capacity models are really models that describe the division, rather than the selection, of attention.

Figure 4.10 shows Kahneman's model. Let's walk our way through it. The rectangular box at the top represents the pool of available attention. A person's arousal level is one thing that can affect this capacity. **Arousal** refers to physiological activation and is reflected in values such as cardiac and respiratory rates. Moderate levels of arousal are assumed to produce the greatest amount of available capacity. This follows from the Yerkes–Dodson law, which states that there is an inverted U-shaped performance function for arousal (Yerkes & Dodson, 1908). Performance is poor for low and high levels of arousal, wherein individuals are presumably too tired or anxious, but it is optimal at intermediate levels.

Miscellaneous determinants of arousal include things such as the amount of sleep a person has had or the amount of food he or she has consumed. If a person has not slept well or has not eaten recently, arousal and, therefore, capacity are assumed to be low. Miscellaneous manifestations of arousal can include overt behaviors such as finger tapping or other nervous habits that result from excessive arousal. Notice, too, that the circle that depicts the evaluation of demands on capacity can itself affect capacity. If there is a greater demand for attention, there may be increased arousal. Knowing that you have a final examination early tomorrow morning might increase the attention you give to studying the night before.

Attentional capacity must now be allocated to the cognitive activities that need it. How is this determined? The ellipse in the center of Figure 4.10 represents the allocation policy. We can think of it as an executive or a decision maker who determines how much attention will be given to a variety of tasks. It is influenced by several factors. First among these are enduring dispositions. An **enduring disposition** is an unalterable bias in how attention is automatically directed toward stimuli that are always relevant, such as a sudden noise or abrupt movement nearby, or hearing one's name from across the room. Stimuli like that automatically draw attention. By contrast, momentary intentions also influence the allocation of attention, but they are not automatic. A **momentary intention** is a conscious decision to pay attention to something. Momentary intentions reflect our specific goals at a given moment in time. A mother driving a car might redirect her attention to her children fighting in the back seat, but she would quickly have to direct her attention back to her primary goal of driving. Notice that the evaluation of demands on capacity and the available capacity also have arrows pointing to allocation, indicating that they, too, exert an influence. The amount of available arousal and the urgency of a task both have an effect on the amount of attention that will go to a particular task.

Figure 4.10 Kahneman's capacity model of attention.



Source: Adapted from Kahneman (1973).

The rectangular box at the bottom of Figure 4.10 contains a representation of possible activities. These are all the cognitive processes or tasks that need attention. The allocation policy that gives attention to them is flexible. For example, equal (rather small) amounts of attention could be given to all the activities, or all available attention could be given to just one activity. Those activities with the greatest demand will be given the most attention. In the driving example given above, the driver's attention was momentarily

taken away from driving—the most demanding activity—and given to the children in the back seat and was then switched back again.

A similar capacity theory for distributing attention to objects in the environment was proposed by neuroscientist Robert Desimone and psychologist John Duncan in 1995, called "biased competition." This framework has become especially influential in the field of visual attention (Desimone & Duncan, 1995). The *bias* component of their theory draws insight from the filter models above, allowing high-level cognitive preferences to redirect attentional resources to parts of the environment that are especially relevant for the task at hand. The *competition* component of the theory draws insight from capacity theories, in that since neural representations for different objects compete against each other, it means that as one representation gets more active, it begins to win the competition and can suppress the other representations.

Evaluating the Capacity Model of Attention

Stephen Reed (2000), in his book *Cognition*, makes some summary comments on attentional theories. He notes that capacity models such as Kahneman's (1973) and Desimone and Duncan's (1995) are not necessarily designed to replace selection models but, rather, to supplement them. Both types of models predict that it will be difficult for an individual to pay attention to two things at once (but for different reasons). According to selection theories, it is because a bottleneck develops, prohibiting the entry of two packages of information at the same time. According to capacity theories, it is because the demands of the two tasks exceed available capacity. Capacity models give us a good overview of the many influences and interdependencies that are involved in the allocating of attention. Selection models provide us with insight into just how far into the perceptual system a cognitive attentional preference might be able to reach for filtering out irrelevant sensory signals.

EVALUATING THE MODEL-BUILDING APPROACH

Our discussion of perceptual and attentional theories so far has demonstrated several things. First, even within a discipline, there is no single widely agreed-on model that explains how a given cognitive process operates. In the case of pattern recognition, we summarized four major models of how it might take place. Second, although these models may vary widely, each model captures some unique aspect of the process better than the others. The feature detection model is better suited to explain errors in recognition, whereas the componential approach (discussed in the Interdisciplinary Crossroads section) can best account for shape constancy. Third, and the most important, we learn more about a process by evaluating all the theories and comparing competing theories. The problems posed by the template matching model show that any model of pattern recognition must extract some invariant feature from the image. That is what is postulated in one form or another by the remaining theories. This cross-theory analysis may not tell us what the features actually are—we see that they differ slightly from model to model—but it does suggest that the process of feature extraction is a key element in the process.

Another major point worth mentioning here is that there is cross-fertilization between model formation and experimentation. In the case of attention, we have seen a constant tug and pull between the models and the experimental evidence. This dynamism led to the formation of better and more improved models. Broadbent's filter theory showed us the basics of how a selective attentional system might operate. Later research, however, proved that information could get through the unattended channel. This led to the formation of Treisman's attenuation model, which by itself could not adequately explain selection that is based on semantic characteristics. This, in turn, led to the development of the Deutsch–Norman memory selection model. The results of later experiments necessitated the introduction of multimode models and the idea of a movable filter, which helped set the stage for Desimone and Duncan's biased competition model. Drawing on this short history, it becomes obvious that we should not put too much faith in any one model at any one time. Instead, we must realize that each model is a point along a developmental continuum, and we must also encourage researchers to improve their models based on the most recent research findings.

SUMMING UP: A REVIEW OF CHAPTER 4

1. In the cognitive approach, the mind tends to be viewed as a computer that represents and processes information.
2. A module is a functionally distinct information-processing component that receives, transforms, and transmits information to other modules, without being influenced by other modules during its processing.
3. Pattern recognition is the ability to identify objects in the environment. The major theories of visual pattern recognition are template matching, feature detection, recognition by components, feature integration, and Marr's computational approach.
4. Marr's computational theory of pattern recognition posits that an image is processed in a number of different stages. A raw primal sketch of the stimulus with light and dark values is generated first. Feature grouping, surface, and layout are processed next in a 2½-D sketch. Object parts are linked and organized in the final 3-D sketch.
5. In template matching, an image of a stimulus is matched to an internally generated representation. If the two match, then recognition occurs. The problem with this process is that it requires different templates for each of the many ways the stimulus can vary.
6. In the feature detection theory of pattern recognition, features such as oriented line segments are extracted from the stimulus input. The number and type of features present are used to diagnose object identity.
7. In feature integration theory, features are extracted preattentively and then combined in a focused attention stage. Preattentive processing is all that is required to locate a target that differs in one way from distractors. Focused attention is needed to locate a target that differs in more than one way from surrounding distractors.
8. Attention may be thought of as a concentrated mental activity that is allocated among

alternative information sources. Attention can be selective, as in focused on one source. It can also be divided between sources or shifted among them.

9. Historically, there have been several major theories of attention. In the filter model, information from one source is allowed to enter conscious awareness. Information from other channels is blocked completely very early in processing. In the attenuation model, some filtered information gets through. In the Deutsch–Norman model, information can be filtered later on the basis of semantic content. The multimode model allows for both early selection based on sensory features and later selection based on meaning.

10. In the capacity models, attention is viewed as a resource that can vary in terms of available capacity, how it is allocated, and how many available activities need it. Kahneman's capacity theory relied on a box-and-arrow diagram for its architecture. Desimone and Duncan's biased competition theory relies on neuroscience.

SUGGESTED READINGS

Fodor, J. A. (1983). *The modularity of mind.* Cambridge: MIT Press.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ: Prentice Hall.

Marr, D. (1982). *Vision.* San Francisco, CA: Freeman.

Nobre, K. & Kastner, S. (Eds.). (2014). *The Oxford handbook of attention*. Oxford, England: Oxford University Press.

Pashler, H. (1998). *Attention.* Hove, England: Psychology Press.

Reisberg, D. (2001). *Cognition: Exploring the science of the mind* (2nd ed.). New York, NY: W. W. Norton.

CHAPTER FIVE

THE COGNITIVE APPROACH II

Memory, Imagery, Concepts, and Problem Solving

TYPES OF MEMORY: HOW DO WE REMEMBER?

In this chapter, we continue our discussion of cognitive theories as we focus on four other prominent cognitive processes: memory, imagery, conceptual representation, and problem solving. In the previous chapter, we looked at processes that may be considered "low level." That is because they focus on information coming in from the environment and how to pay attention to it. In the present chapter, we examine the so-called high-level processes because they focus on how information is represented and processed more internally with somewhat less reliance on the immediate environment. We will describe the major theoretical approaches to each of these major processes as well as some of the classic experiments that led to their development. There has been an immense amount of research on these and other topics in cognitive psychology, and by necessity, we must leave out many interesting and important findings. As in previous chapters, we refer interested readers to the works in the "Suggested Readings" section if they wish to learn more of these findings.

Memory, very generally defined, is the capacity to retain information over time. Memory is, of course, very important to any information-processing system, animal, or machine, because it underlies the ability to learn. Any system incapable of learning from its mistakes would soon perish or become obsolete in a dynamic, real-world environment. Memory allows us to store our past experiences and draw on them. In this way, we can deal with new situations that bear similarity to older situations and not have to solve each new problem from scratch.

Any discussion of memory must begin with the notion that there is no single type of memory. Cognitive research in this area has

Learning Objectives

After reading this chapter, you will be able to:

1. Define sensory, working, and long-term memory.
2. Assess the modal memory model's three modules for memory.
3. Discuss the evidence for Kosslyn and Schwarz's theory of visual imagery.
4. Contrast defining features and typical features in conceptual representations.
5. Define a perceptual simulation in Barsalou's theory of concepts.
6. Sketch the concept of a problem space, with initial, intermediate, and final states.

led to a rather consistent finding: the existence of functionally distinct memory systems. Models of memory specify how these different memory systems interact. Before we can understand this interaction, we must lay out some of the characteristics of these individual memory systems. They include duration, which refers to how long information remains viable in a memory system; capacity, which refers to how much information the memory system can hold; and **coding**, which refers to the particular type of information the system contains. In the first part of this chapter, we will summarize these characteristics. Following this, we will describe formal memory models.

Sensory Memory

Sensory memory is a repository for incoming sensory information. Raw, unanalyzed data that are derived from the senses are held here very briefly. The purpose of sensory memory is to maintain the representation of a stimulus long enough for it to be recognized. Although you may have glanced at a visual scene for only 100 milliseconds, a representation of it is still preserved in sensory memory. This gives the information a chance to be operated on by selection and pattern recognition mechanisms.

There are different forms of sensory memory—one for each of the five senses. Each of these forms has different characteristics. **Iconic memory** is visual sensory memory. It holds a brief "snapshot" of what you have just looked at. Iconic memory has a very short duration; it lasts only about 250 to 300 milliseconds (Averbach & Sperling, 1961; Sperling, 1960). **Echoic memory** is an auditory sensory store. You can think of it as an "echo" of what you have just heard. It lasts considerably longer than iconic memory, on the order of several seconds or more (Darwin, Turvey, & Crowder, 1972). The coding or representation of information in sensory memory thus varies with the modality. Iconic memory stores visual representations; echoic memory stores auditory ones. More research needs to be done on sensory memory for the other senses, including olfactory (smell), gustatory (taste), and tactile (touch). In each case, we are dealing with a different type of representation, one corresponding to each modality. So we see that the issue of representation gets complicated; we can have multiple representations for each sensory modality.

George Sperling first began investigating iconic memory in the 1960s. He presented participants with a 50-millisecond display of English alphabet letters arranged in a four-column by three-row letter array. On a given trial, a display might look something like the following:

R G C P

L X N F

S B J Q

There were two conditions in this study (Sperling, 1963). In the **whole-report condition**, the participant's task was to remember as many of the letters as possible. Most participants recalled only four or five. Sperling had a hunch that they were actually remembering more than this. He suspected that the procedure was testing not for

what is available in the icon (in this case, the entire display of letters) but for what we pay attention to or process after the fact. Four or five individual items might be what is remembered but not what is actually present in the sensory icon. To investigate this, he developed a **partial-report condition**. After being shown the letter display for a mere 50 milliseconds, participants immediately heard one of three tones. A high-pitched tone cued the top row, indicating that the participants would be expected to report on that row; a medium-pitched tone prompted the middle row; and a low-pitched tone cued the bottom row. The results then showed that participants were able to recall at least three of the four letters in the cued row. Since the visual display had disappeared by the time they heard the tone, this shows that there was some form of visual icon of nearly all 12 letters available in memory to which the row-cue was applied. However, when Sperling introduced a 1-second delay between the brief presentation of the letter display and the cue tone, participants were only able to recall one or two of the letters in the cued row.

We learn from this research that the capacity of the visual icon—how much it can hold—is at least 12 items. Subsequent research has shown that iconic memory can, in fact, hold much more than this. Its capacity is essentially unlimited. Everything that can be taken in during a glance—all the information in the visual field—can be stored there. But the integrity of this information is short-lived, not lasting more than 1 second. The image is clearly represented early on but then fades quite rapidly. The loss of information over time in memory is known as **decay**.

More recent research on iconic memory has revealed various factors that can influence performance. Becker, Pashler, and Anstis (2000) presented a visual array of items that could be followed by either a blank interstimulus interval (ISI) or an ISI with a cue providing the location of a changed item. Change detection performance was poor in the blank ISI condition but improved significantly in the cued condition. They hypothesize that the blank interval overwrites or erases the icon formed in response to the first array, whereas attention to the cue insulates the icon against overwriting. This erasing process is called **masking** and is manipulated as an independent variable in many different types of perceptual studies.

In another work, Thomas and Irwin (2006) presented observers with a letter array using the partial-report method. When participants were instructed to blink following the offset of the target array and the onset of the cue, they made more mislocation errors. They conclude that blinking interferes with the binding of object identity and object position in iconic memory and named this phenomenon *cognitive blink suppression*. Blinks were found to reduce neural activation in area VI (primary visual cortex), which may disturb formation of the icon.

What other brain areas might underlie iconic memory formation? Keysers, Xiao, Foldiak, and Perrett (2005) presented brief visual stimuli with and without interstimulus gaps to humans and monkeys. For an ISI of 93 milliseconds, they found continued neuronal activation in monkeys' temporal cortices as if the stimulus was still present on the screen. Ruff, Kristjansson, and Driver (2007) used six items in their displays, three in each hemifield (left or right visual field). Their auditory cues indicated which side of the display to report. In the precue condition, which measures selective spatial attention, the cue came before the array. In the postcue condition, measuring iconic memory, it

came after. Using functional magnetic resonance imaging techniques, they found that both types of cues led to contralateral (opposite side) activation in the lateral occipital cortex. This suggests that read-off of iconic memory and spatial attention share some underlying neural mechanisms. They also found bilateral activation in the frontoparietal network for each cue type. However, postcues triggered more activity in the right middle frontal gyrus.

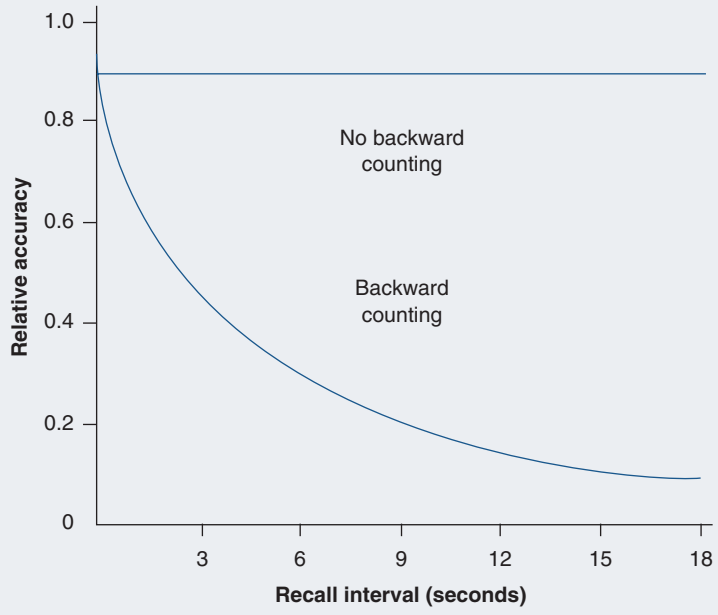
Working Memory

Working memory is also sometimes known as short-term memory. Researchers sometimes use these terms interchangeably, as will we throughout this text. As the term *short-term memory* suggests, information is briefly stored there. However, the duration of items residing in working memory is, as we will see, much longer than that of items residing in sensory memory. Working memory retains less information; its capacity is unlike the unlimited capacity of the visual icon. Working memory is limited to storing just a small number of items. Whereas information in the different sensory stores is specific to modality, coding in working memory can be acoustic, semantic, or visual. We will have more to say about how the different types of codes are put to use when we discuss Alan Baddeley's model of working memory later in this chapter.

It is helpful to use the analogy of a workbench when thinking about working memory. A workbench is a space where one can construct something using parts and tools. For example, if one wanted to build a spice rack, he or she would need several pieces of wood, some nails, and varnish. In addition to these, he or she would need saws—to cut the wood to size, a hammer, and a paintbrush. With these items on the bench, one could set about creating the final product. Working memory is the mental equivalent of this workbench. It is a place where data (the parts) can be temporarily stored so that they can be operated on by cognitive processes (the tools). In this sense, working memory is the site where conscious thinking takes place. It is here that you remember a phone number that you've just been told, figure out how to navigate your way around a new city, or solve an arithmetic problem.

In a classic study, Peterson and Peterson (1959) demonstrated the duration of items that reside in short-term memory. They presented participants with three letters to be remembered. A tone was sounded following the presentation of the items. In separate trials, the tone was sounded at varying time intervals. Sometimes, it would go off immediately; during other trials, there would be a substantial delay. The participants were instructed to recall the items when they heard the sounding of the tone. In this condition, all participants were able to remember the items correctly, no matter the length of the delay. In a second condition, presentation of a three-digit number followed presentation of the items. Participants were instructed to count backward in threes from this number. If the number was 796, they would have to count out 793, 790, 787, and so on. A tone was sounded, again at varying intervals, cueing recall of the items. At this point, the results were quite different. The ability to remember the items deteriorated rapidly as the period of delay increased. After an 18-second recall interval, the accuracy of recall had dropped to just 5%, meaning that participants could recall only 5% of the items correctly. Figure 5.1 shows idealized results for both conditions of such a study.

Figure 5.1 **Idealized results of a memory study in which participants are asked to remember an item after a given recall interval. Having to count backward after presentation of the item interferes with rehearsal.**



You may have guessed what's happening here. In the first condition, participants were able to mentally rehearse the letters. **Rehearsal** is the mental repetition or mental "practicing" of some to-be-learned material. It is usually manifested as implicit speech, or "talking to oneself"—what you do when you need to remember somebody's phone number but can't write it down. Rehearsal refreshes items in short-term memory; it keeps them active and prevents them from decaying. In the second condition, participants were prevented from rehearsing because of the backward counting. Here, the items underwent decay. Using this paradigm, Brown (1958) and Peterson and Peterson (1959) were able to determine that the duration of short-term memory was about 18 seconds. If you are interested in some of the other processes that can act on information in working memory, go to the next section, where you will see that—in addition to rehearsal—items can be scanned.

Our next discussion moves from *duration* of working memory to its *capacity*. How much can working memory hold? It is easy to devise an experiment that tests this question. It would involve presenting participants with a series of lists, each containing a different number of items. For example, we could present a list of four digits, then a five-digit list, a six-digit list, and so on. If there were a limit to how many digits could be held

in working memory, we would see a dramatic drop-off in retention as we arrived at that limit. Many early experiments of this sort demonstrated that this was indeed the case. On average, individuals can retain about seven items, give or take two, commonly denoted as 7 ± 2. This limit has come to be called the "magical number seven" (Miller, 1956).

Therefore, short-term memory has a limited capacity. We can hold only about seven items in it at one time. But what exactly is an "item"? Is it a number? A letter? A word? To understand this, let's start with a demonstration. Glance quickly at the letters below and then close your eyes to see how many you can remember.

N F L C B S I R A M T V

Did you have a hard time? If you weren't able to recall all the letters easily, take a closer look. You may at this point notice that the letters cluster in groups of three—each group is a popular acronym. "NFL" stands for National Football League, "CBS" for Columbia Broadcasting System, and so on. If you had noticed these groupings, it would have been easy for you to remember all the letters. This is because you would then be remembering only four groups of three letters and not 12 individual letters. A grouping of items into a single meaningful whole in short-term memory is known as **chunking**. We can increase the total amount of information that is contained in short-term memory by chunking: The bigger the chunk, the greater the capacity.

Let us turn now to the issue of coding. How is information coded or represented in working memory? Research in this area shows that information can be coded in a variety of formats. Conrad (1964) presented letter strings to participants and then asked them to perform immediate recall. He found that although the letters had been presented visually, the pattern of errors was based on the sounds of the letters. For example, participants would mistake an *A* for a *K* or an *E* for a *P*. This suggests that the letters were being converted to an **acoustic code**, one based on the sounds of the items.

Studies by Wickens (1972) and others have demonstrated the existence of an alternate code that is part of working memory. To make sense of their experiments, we must first differentiate between two types of interference. In **proactive interference**, information that is learned earlier interferes with information learned later. In **retroactive interference**, information that is learned later interferes with information learned earlier. The more closely related items are with respect to meaning, the greater this interference. Wickens presented words to subjects along the lines of the tasks that were given to participants in the Brown and Peterson experiments. He found that after they had studied words that belonged to a single category, there was proactive interference, which engendered a steady decline in recall over the first few trials. For example, the word *apple*, learned earlier, would interfere with recall of the word *orange*, learned later. But in a final trial, during which the semantic categories of the words the participants were asked to remember were sometimes switched—in one condition, from fruits to flowers—there was an increase in recall. In this instance, the word *apple*, learned earlier, would interfere less with the word *tulip*, as they belong to separate semantic categories. These results showed that the words were represented not acoustically but, instead, on the basis of their inherent meanings—the code was a **semantic code**.

Additional studies showed that information in working memory may be represented in yet a third way. A **visual code** is one that preserves spatial characteristics. In studies of the mental rotation of objects, participants were presented with a visual pattern and were asked to compare it with another that was either a different pattern or the same pattern rotated by a certain angle. The plot of reaction time in these studies showed that subjects apparently formed a visual representation of the pattern and rotated it mentally to enable them to perform the comparisons (Cooper & Shepard, 1973; Shepard & Metzler, 1971). The representation seems to be visual because the greater the angular difference between the two patterns being compared, the longer it takes to respond—exactly what one would expect if a real spatial object were rotated. We will describe these experiments and the controversy that surrounds them in greater detail later in this chapter.

Scanning Items in Working Memory

You go to the grocery store to buy some food for dinner. Because you need only a few items, you memorize the items rather than writing out a list. After buying rice and broccoli, you pause to consider what else it is you have to purchase. What mental process occurs when we perform this sort of task? The process in question is called "memory search," and it is one of the many operations carried out in working memory. In this section, we describe a classic experiment from cognitive psychology that suggests how such a mechanism might work. Please note that search in working memory is not the same process as the visual search we discussed in the previous chapter.

Search, in an abstract informational sense, involves looking for a target among a set of items. Each of the items must be represented mentally, as must the target. Then, a representation of the target must be compared against each of the represented items in the set in the expectation that a match will be made. But now a question arises. This search process can proceed in two ways. If it is a **serial memory search**, then comparisons take place one at a time, with a participant perhaps starting at the beginning of the list and proceeding through until the end. If this is the case, then search times should be long, as each comparison takes time. If it is a **parallel memory search**, then the target can be compared with all the items at once. Assuming this to be the case, search times should be short, as only one comparison is being made.

To find out whether the search was serial or parallel, Sternberg (1966; see also Sternberg, 2016) asked participants in his study to memorize a set of items, such as a series of digits. The size of the set varied between one and six digits—always below the 7 ± 2 limit. Sternberg then presented a target digit. The participants had to decide whether the target was a member of the memory set or not. For example, in one trial that used a set size of four, the participants might have had to memorize the digits 4, 9, 0, and 2. They were then given a target, the number 9. Following the presentation of the target, they were to respond as quickly as possible, pressing one button if the target was present in the set (a positive trial) and pressing another if it was absent (a negative trial).

Reaction time (the time it took to press the buttons) was analyzed and plotted against set size for both positive and negative trials. Sternberg made two predictions. If the search was serial, he should obtain an increasing function with longer search times for larger set sizes. This follows, because if each comparison takes a given amount of

time, a greater number of comparisons will require more time. If, on the other hand, the search was parallel, he should obtain a flat function over set size, indicating the shorter time it takes to compare the target with all items in the set at once, no matter how many items are in the set.

Sternberg made yet another prediction. He speculated that the search, if it was serial, could be self-terminating or exhaustive. In a **self-terminating search**, the search process stops as soon as a positive match occurs. Because the target can be located anywhere in the set, the search times will be variable. Sometimes, if the target matches an item at the beginning of the list, the search will end quickly. On other occasions, when the target matches an item that is near the end of the list, it will take longer. This effect manifests itself as lesser slopes in the search function for positive trials, in which a target is present in the set, but as steeper slopes for negative trials, in which scanning must always continue to the end. In an **exhaustive search**, comparisons are made all the way to the end of the list, which would be necessary if the target was not on the list. Curiously, positive trials behaved as though they involved an exhaustive search as well (see Figure 5.2). If positive trials used a self-terminating search, then the slope of their reaction times as a function of set size should have been half as steep as the slope for negative trials. However, the slopes were the same for positive trials and negative trials.

Data showed that the memory scanning was serial and exhaustive for both positive and negative trials. The search function increased linearly with set size as steeply for positive trials as for negative trials. At first blush, this result seems counterintuitive and inefficient. Why should you continue looking for something you've already found? There are several possibilities. Exhaustive search could serve as a "double check," a way to ensure that the target isn't duplicated later in the list. Or it could be that once the search is initiated, it proceeds automatically and cannot be stopped. The latter possibility is interesting because it implies that the search is only partially under voluntary control. The central executive in the working memory model may issue a command to the articulatory loop to begin a search, but once this command is issued, it cannot be further regulated.

Long-Term Memory

Obviously, individuals are capable of remembering information for longer than just a few seconds. Many of us can recall quite vividly events from our childhoods or facts learned decades ago in school. We must, therefore, have another memory system, one that is capable of storing data for months or years.

It turns out, as is the case for memory in general, that long-term memory has several distinct types. **Procedural memory** holds procedural knowledge. It is memory for skill, is demonstrated by doing, and arises without conscious recall. Knowing how to ride a bicycle is a good example of this memory type. A person who knows how to ride a bike can demonstrate that he or she has this ability only by actually doing it. The subconscious nature of this memory becomes evident when we learn how to perform some skill—such as playing the piano—forget how, but then show improvement when we attempt to perform the action at a later date. Procedural memory also sometimes goes by the name of implicit memory.

Figure 5.2 **The results for the Sternberg experiment showed that memory search was serial and exhaustive. Note that negative trials (target absent) and positive trials (target present) have the same slope, suggesting that they both involve serial exhaustive search.**



Source: Sternberg (1966).

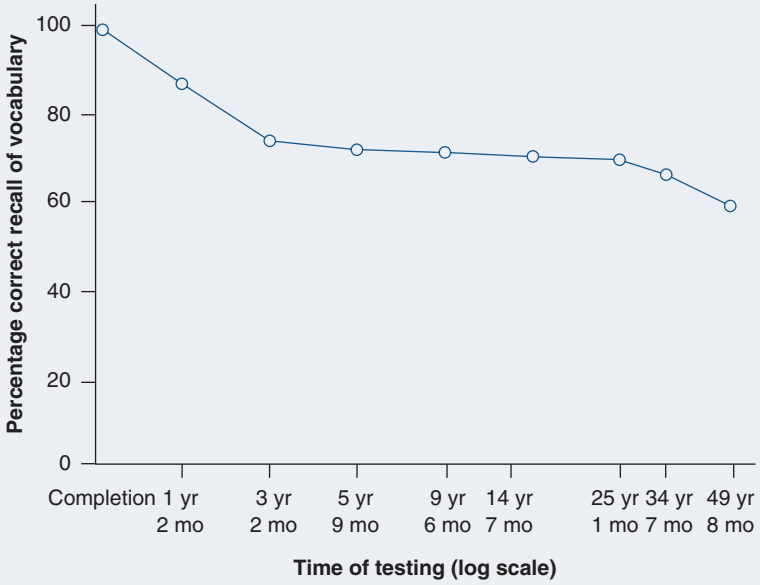
Declarative memory contains declarative knowledge. It is memory for facts and events, is demonstrated by speaking, and arises with conscious recall. There are two types of declarative memory. Knowing that Thomas Jefferson wrote the Declaration of Independence is an example of the first type, called semantic memory. **Semantic memory** is knowledge of facts and general knowledge of the sort learned in school. The second type is episodic memory. **Episodic memory** contains episodes or personally experienced events—for example, what you did on your last birthday. We are usually consciously aware of declarative information, which is sometimes referred to as explicit memory.

A study by Bahrick (1984) shows us how long information in semantic memory lasts. Bahrick studied a group of people who studied Spanish early in life, usually during the

high school or college years. They were tested at various times throughout their lives with respect to how much they remembered from their Spanish classes. When recall for various types of information was plotted as a function of time, three distinct memory stages were discovered (see Figure 5.3). In the first stage, stretching from 3 to 6 years after the classes, there was an initial, rather dramatic, loss of information. In the second stage, stretching from 6 up to 30 years after initial learning, there was no further loss—the amounts remembered stayed relatively constant. In the third stage, stretching from 30 to 35 years later, there was another loss of information, but only a slight one.

We can draw conclusions from each of the three stages. First, information in long-term storage, if not used or rehearsed, decays fairly rapidly over the first few years. Second, there is some residual amount of learned information that remains intact over a long period of time. Third, there appears to be a slight additional loss of information that occurs later in life. This loss could be due to the kind of general decline in cognitive function that accompanies aging. The basic shape of this memory function stayed the same regardless of the levels of training or the grades the participants had received. Better initial learning, as measured by these indices, resulted in an elevation of the function such that overall performance was better but the shape of the function remained unchanged.

Figure 5.3 Percentage of correct recall for Spanish learned in school, at different time periods after original learning.



Source: Adapted from Bahrick (1984).

Now back to capacity. How much can be retained in semantic long-term memory? It has been proposed that we remember virtually everything we've ever experienced in our entire lives but simply have difficulty recalling it. Therefore, although information may get into long-term memory and stay there without its being lost, our inability to remember it could be due to a failure in "getting it out." One researcher estimates that the average adult has about a billion bits of information in memory and a storage capacity that is perhaps one thousand to one million times greater than that (Landauer, 1986). However, we must be skeptical about such estimates, since the inability to recall an item in a memory test can be due to either retrieval failure or decay.

With respect to coding, the current belief is that information in long-term memory is represented in various formats. Implicit memories may be stored in a format that is roughly equivalent to production rules—formalized if–then statements that match a sensory input to a motor output. One proposed neural locus for production rules like this is in the cerebellum, a part of the brain that mediates motor learning. Refer back to the introductory chapter for more on production rules. Explicit memories may be stored as networks of connected nodes, forming distributed representations of the facts and events. The distribution of these nodes is most likely throughout the entire cortex and other diffuse brain areas. See the network approach chapter for more on these networks.

MEMORY MODELS

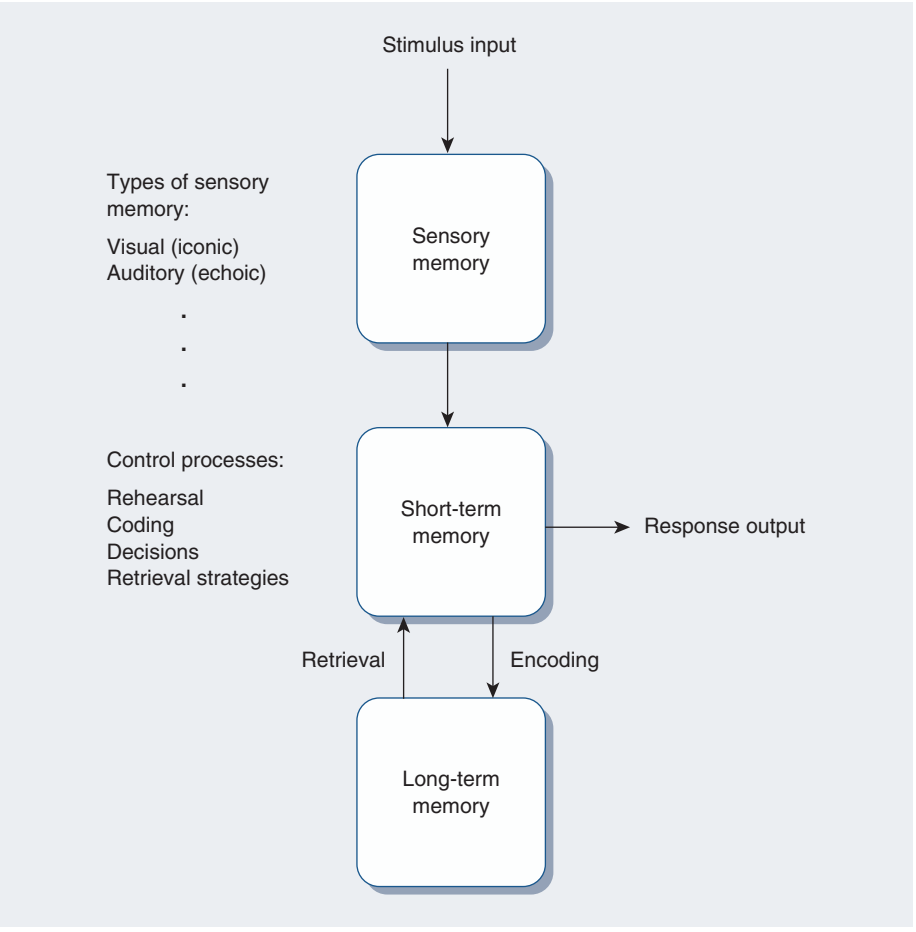
We turn now to summarizing some of the most influential theories of information processing that attempt to describe memory. These theories specify interactions between all the different memory types we've seen so far. Alternatively, some of these theories describe the interactions between components of a single memory system. In each case, the interactions are conceptualized through the use of process models that show how information is transformed and shuttled back and forth among the successive stages of processing.

The Modal Model

The **modal memory model** was the first model to provide a general overview of how information is processed in each of the different memory types (Atkinson & Shiffrin, 1971). Figure 5.4 shows the stages of the modal model. To start, stimulus information from the outside world first enters sensory memory, where, as we have seen, it is very briefly held. From there, the information is transferred to short-term or working memory, where a wide variety of computations can be performed. These include rehearsal and recoding. Following this, to-be-remembered information is passed to long-term memory. **Encoding** is the name of the process by which information is taken into long-term memory and converted into a usable form. A related process, discussed at length in the neuroscience chapter, is consolidation. Consolidation is another process by which information is transferred from short- to long-term memory. Consolidation "strengthens" information so that it lasts longer.

Once in long-term memory, information may not be used immediately. We say that the information is in a state of **storage**. Stored information is represented but not

Figure 5.4 The Atkinson and Shiffrin (1971) modal memory model.



available for immediate use. As alluded to above, this information may sit in limbo for a very long time. When information that resides within long-term memory is needed, a retrieval process takes place. **Retrieval** is the act of accessing needed data and making them available for use. In the modal model, retrieval is shown as the transfer of the information from long-term to working memory, where it can then be operated on.

Evaluating the Modal Model

The modal model provides us with a good summary of memory function and serves as a useful starting point. However, it relies heavily on a metaphor of treating memories as though they were items that are stored in a container. Rather than pretending that

memory is a container, one can instead examine the properties of neural networks (see Chapters 6 and 7) to find that sensory, short-term, and long-term memory capabilities can emerge naturally from how neurons interact with one another. For sensory memory, this can result simply from a residual activation pattern that reverberates among the activated neurons for several hundred milliseconds. When a stimulus goes away, the neurons that were responding to it cannot turn themselves off instantaneously; they exhibit a brief residual activation that can serve as a sensory memory (Keysers et al., 2005; Lu & Sperling, 2003; Ruff et al., 2007). For short-term working memory, positive feedback connections among a set of neurons can maintain a given pattern of activation for many seconds or even minutes (e.g., Durstewitz, Seamans, & Sejnowski, 2000). For long-term memory, knowledge is encoded via a learning algorithm that subtly rewires the neural connectivity of the network (e.g., Hebb, 1949; Rumelhart, McClelland, & the PDP Research Group, 1986). New skills and facts are acquired long term not by placing symbols into a container but instead by changing the synaptic strengths of neural connections. (Of course, in order to accommodate the distinctions between different types of long-term memory, such as semantic memory and episodic memory, additional parameters or subcomponents would need to be included in such a network.) Thus, rather than treating these three types of memory, with their very different time scales, as involving separate boxes in the mind that are lined up in sequence, they may instead be described as three different capabilities that a complex neural network naturally exhibits. The modal model was developed very early in the course of cognitive psychology and fails to specify many of the nuances of memory structure and function. Upgrading the model would, at the very least, require the postulation of additional memory subtypes and the insertion of a number of additional processing steps. For instance, the modal model fails to acknowledge the large number of information-processing operations that can be performed in working memory. More recent models of working memory satisfy this and are discussed later in this chapter.

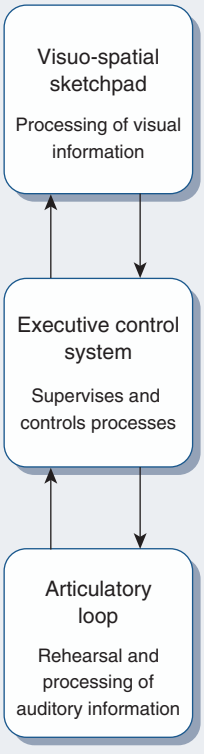
The Working Memory Model

Baddeley (1986, 1992, 2012) has formulated a detailed model for the components and processes of working memory. These are shown in Figure 5.5. In Baddeley's scheme, working memory is composed of three systems. He calls the primary unit the **executive control system**. The job of this system is to initiate and control ongoing processes. Like the captain of a large ship, this system issues orders and directs subsidiary crews. Some of its activities are reasoning, language comprehension, information transfer to long-term memory via rehearsal and chunking, and information retrieval, also from long-term storage.

The second major component in Baddeley's model is the **articulatory loop**, sometimes called the phonological loop. This is the place where speech- and sound-related information is rehearsed. Information such as a telephone number that you want to remember and actively maintain for immediate recall is passed to the articulatory loop to refresh it. A third system, called the **visuospatial sketchpad**, is specialized for the processing of visual information. It is here that visual imagery tasks such as mental rotation or visual search are performed. If you were to form a mental image of a cat with your eyes shut, the image would be represented in the visuospatial sketchpad.

Both the loop and the sketchpad are "slave systems" to the central executive. They carry out processes such as rehearsal and image formation. Other processes, such as reasoning and language, are the responsibility of the executive. Notice also that these two slave systems are domain specific: One is devoted to the processing of acoustic information only, the other to the processing of visual information only. This means that each one operates on a different code—the loop on an acoustic code, the sketchpad on a visual code. Tests of working memory have shown that each slave system relies on its own pool of attentional resources (Baddeley & Hitch, 1974; Logie, Zucco, & Baddeley, 1990). If the tasks that one of these systems is performing are simple and don't require much in the way of attentional resources, then working memory performance is unchanged. If one of the systems is given a demanding task, though, it either fails to complete the task or draws on the attentional resources of the executive, which results in impaired working memory performance.

Figure 5.5 The different components of the working memory model.



Source: Baddeley (1986).

Evaluating the Working Memory Model

Baddeley's working memory model is in essence a reformulation of the traditional idea of short-term memory. This idea would have us believe that short-term memory is a relatively inactive location where information is either rehearsed or decays. Baddeley extended our concept of this memory system to encompass a self-regulating collection of processors that perform imagery, reasoning, and language tasks. Recent studies have shown that specific brain regions correspond to the different components of the Baddeley model (Smith, 2000; Smith & Jonides, 1999). Verbal storage, rehearsal, and executive tasks were found to activate a number of regions in the parietal and frontal lobes of the left hemisphere. In contrast, visual processing activated regions in the occipital, parietal, and frontal lobes of the right hemisphere. Much work in the area of uncovering exactly which types of neural activity underlie the specific operations of working memory remains to be done.

VISUAL IMAGERY: HOW DO WE IMAGINE?

How many windows are there in your living room? What is the shape of a German shepherd's ear? To answer these questions, you may have resorted to the use of visual imagery. A **visual image** is a mental representation of an object or a scene that preserves metric spatial information. Visual images are, therefore, isomorphic to their referents. They preserve spatial characteristics of the objects they represent. An example would include distance. If it takes longer to scan what is between two parts of a physical object because the distance between those parts is relatively greater, then it should also take longer to scan between those parts in the mental image representation.

Isomorphisms do not imply that images have the same physical properties as the objects themselves. An imagined banana is, of course, not really yellow or curved. But the representation of the banana in the image allows it to be treated as if it were a real banana. That is, it may be *mentally* picked up, peeled, or squashed. The image allows us to perform the same sorts of operations on the "banana" in our heads as we would on the actual banana with our hands. Although images in many ways seem to follow the same laws as real objects, they are by no means constrained by such laws. We could, if we wanted, imagine a flying banana or a banana singing the blues!

A defining hallmark of imagery is that, unlike perception, it is not driven by sensory input from the outside world. We can form a visual image of a car with our eyes completely shut. Imagery can be driven by information coming from spoken linguistic input, an external auditory source, or from memory, an internal information source.

It is important to note that *imagery* in a more general sense refers to the mental representation of something and need not be restricted to the visual domain. It is possible to have auditory, or olfactory, imagery as well. For example, one can imagine the melody to the song "Yankee Doodle Dandy" or imagine the smell of a steak being grilled over an open fire. As was the case with perception, most of the research that has been conducted in this area—as well as theorizing on this subject—has focused on the visual, so we'll discuss only visual imagery here.

The Kosslyn and Schwartz Theory of Visual Imagery

In this section, we devote much of our attention to an early and influential theory of imagery (Kosslyn, 1980; Kosslyn & Schwartz, 1977). It is a functional theory, meaning that it describes what the brain does with images. It is also a general theory; it attempts to account for the diverse operations that are involved in image processing, as opposed to it being a specific theory that focuses only on one type of operation. Mental rotation, described in this section, would be an example of a more specific theory. The Kosslyn and Schwartz theory of visual imagery was elaborated in enough detail that it has been implemented as a computer program. In describing this theory, Kosslyn differentiates between the structures that are involved and the processes that act on them. In keeping with this, we will describe the structures first and the processes second.

Image Structures

Kosslyn posits two kinds of image structures. A **surface representation** is a quasi-pictorial representation that occurs in a spatial medium. It depicts an object or a scene and underlies our experience of imagery. But what is this medium? He describes the medium as a **visual buffer**. This buffer is a surface matrix that consists of an array of points, similar to what makes up an image on a TV screen. The visual buffer functions as a coordinate space and has a limited extent and a specific shape. It also has a grain, or limited resolution (Kosslyn, 1975). The resolution is highest at the center and decreases toward the periphery (Finke & Kosslyn, 1980).

Representations in the buffer are short-lived. Much like an image on a computer or TV screen, a visual buffer representation fades away if it is not refreshed. If an object with many parts is imagined, the parts generated first may decay before those generated later. This can explain why images of objects that are part of complex scenes are more degraded than images of objects in simple contexts (Kosslyn, 1975).

According to the Kosslyn theory, certain characteristics of the surface image are unique to analog representations. The images exhibit features not present in propositional representations. To start, the size of the image, the shape of the image, its orientation, and the location of information in the image do not exist independently. When the value of one is specified, values of the others must also be specified. If we were imagining a fish in different ways—for example, as a big or a small fish, a fat or a thin fish, a right-side-up or an upside-down fish, or a fish to the left or to the right—then no part of the fish in the image is guaranteed to stay the same after the transformation. Images in this sense are holistic; a change in the image produces changes across its entire extent. In addition, any part or piece of an image is a representation of a part or a piece of the actual object being represented, the referent. The part of an image of a fish that corresponds to the fin would represent that fin on the real fish.

What about the content of such images? Images are about objects. They are viewer centered and represented according to the perspective of the observer who is viewing them. An image can represent an object directly, as a photograph does, or indirectly, as a diagram or schematic does. Content is determined not just by the image itself but also by how the image is interpreted. The meaning of an image, therefore, does not lie entirely within the image itself but also in how that image is processed by an interpretive device.

Deep representations constitute the second type of image structure. They consist of information in long-term memory that is used to generate the surface representations. Kosslyn describes two classes. **Literal encodings** contain lists of coordinates that detail the placement of points in the surface matrix such that the represented object is depicted. Some evidence suggests that they originate in the right hemisphere. **Propositional encodings** are abstract, language-like representations, similar to declarative statements. They contain information about an object's parts, the locations of these parts, and their sizes. They are believed to lie in the left hemisphere. It helps to think of these encodings in the following way: Imagine that you receive a chair that requires assembly from an office supply store. The chair would come with a list of parts, such as the base, legs, and back, as well as the various nuts and bolts needed to fasten the parts together. This list of parts corresponds to the propositional encodings. The chair would also come with a set of instructions specifying how to connect the parts. It would tell you which bolt to use to fasten Part A to Part B. This instruction list is analogous to the literal encodings.

Image Processes

We have already reviewed the empirical evidence that supports each of the three types of image processes. **Image generation** occurs when the encodings in long-term memory are used to form an image in the visual buffer. Kosslyn envisions this process as a set of distinct operations, each labeled with a distinct command in the computer simulation. A PICTURE command converts the information that has been received from the encodings, FIND locates the position for the placement of a new part, PUT adjusts the size of the imagined part, and IMAGE coordinates other commands and performs operations such as determining the appropriate level of detail.

One question concerning image generation is whether images are formed all at once or bit by bit. The evidence suggests that they are formed bit by bit (Kosslyn, 1980). Images that have more parts take longer to create. The same is true for images that have more detail. Data from another study show that people who are asked to create images have some control over how complete and detailed their images will be; depending on task demands and individual preferences, images can be sketchy or elaborate (Reisberg, Culver, Heuer, & Fischman, 1986). This result is in accord with the idea that images are formed gradually rather than all at once.

Image inspection occurs when we are asked some question about an image—for example, "Is a pig's tail lower than its snout?" Inspection is like looking at the image with the proverbial "mind's eye." It also consists of a number of distinct processes, each labeled with a different command in the computer simulation. RESOLUTION determines if the image is at the right scale and calls on ZOOM or PAN to expand or contract the image. SCAN, as its name suggests, scans between locations within the image.

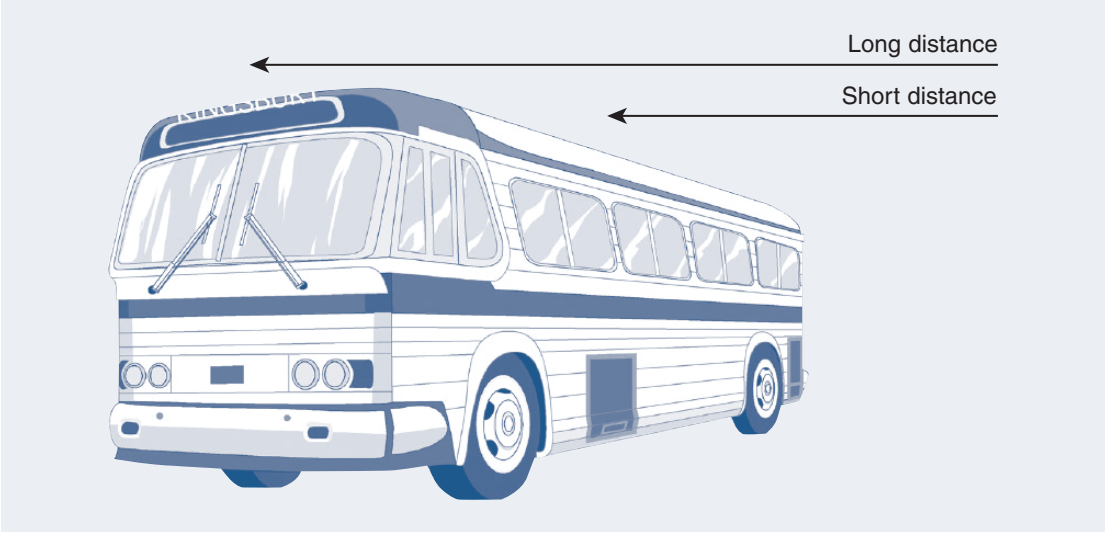
An experiment by Kosslyn (1975) suggests that we can zoom in on or magnify images as part of the inspection process. He asked participants to imagine a large object next to a small one. For instance, subjects were asked to imagine an elephant standing next to a rabbit. They were then asked to respond to questions—for example, "Does the rabbit have a tail?" Kosslyn found that, under these conditions, participants took longer to respond to questions about the rabbit. His explanation was that the rabbit was relatively small in the image and, therefore, difficult to resolve. Kosslyn

inferred that participants zoomed in on or magnified the part of the image in question to respond.

Another study, by Kosslyn, Ball, and Reiser (1978), showed that images may be scanned or "looked across" in the same way that real objects are. To get a sense of what participants were asked to do in this study, try the exercise depicted in Figure 5.6. The participants in the study were asked to memorize a map of a fictitious island. They were then instructed to form an image of the entire island and to focus on one of the objects on it. The name of a second object was announced, and participants were asked to scan their images by imagining a black speck moving in a straight line from the first object to the second. When the black speck arrived at the second object, participants pushed a button, stopping a timer. When reaction times were plotted as a function of map distance, a linear relationship was revealed. Participants took longer to respond when the distances between objects on the island were greater. From this result, it was inferred that images do preserve spatial extent. Figure 5.7 shows the pattern of the results that have been obtained in experiments like this one.

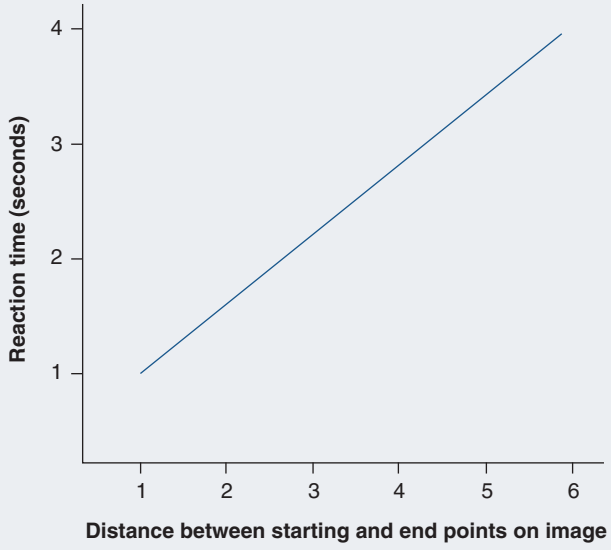
Image transformation refers to an operation performed on an image. Kosslyn describes two transformation modes—shift transformations and blink transformations. Most image transformations are shift transformations, in which the image is altered incrementally. These changes can be seen as continuous changes in which the image

Figure 5.6 Study this picture of a bus for 1 minute. Then close your eyes and form a visual image of it, focusing on the back of the bus. Answer these two questions: Is the luggage compartment open? How many headlights are there? Which question took longer to answer? Why?



Source: Deneba Clip Art.

Figure 5.7 Idealized results from a visual image study showing time needed to mentally scan from one point of an image to another as a function of the distance between objects.



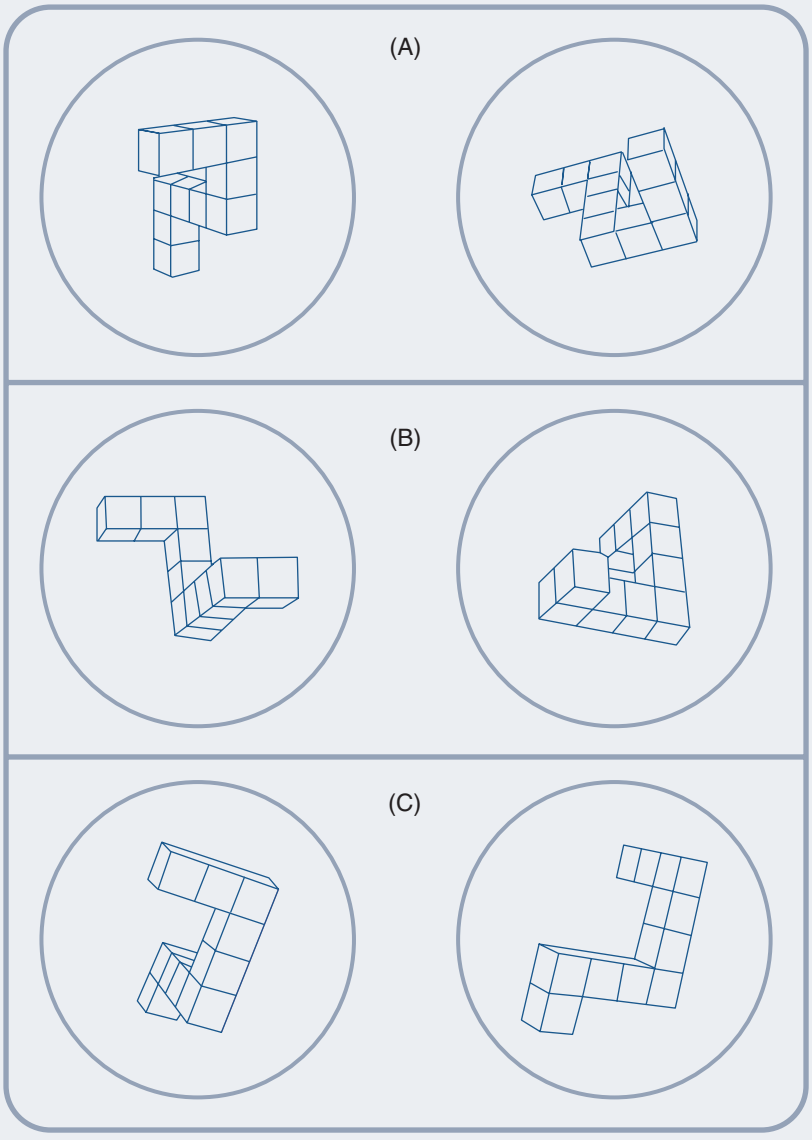
passes through intermediate stages along a trajectory of size, orientation, or location. Blink transformations, in contrast, alter the entire image at once and can be used to re-form a complete image after it has decayed in the buffer.

A classic experiment in cognitive psychology calls attention to an example of a shift transformation. Shepard and Metzler (1971) presented participants with drawings of pairs of three-dimensional objects (see Figure 5.8). Sometimes, the paired objects were different. But sometimes, they were the same, in which case the two objects were rotated versions of each other. The observers were asked to evaluate the pairs and to respond by pulling one lever with their right hand, if the objects were the same, and another lever with their left hand, if they were different. The response times were then plotted as a function of the angle of rotation (see Figure 5.9). The plot demonstrates that the greater the angular difference between the shapes, the longer it took to respond. Shepard and Metzler (1971) concluded that participants had formed a visual image of an object that was based on one of the objects in the drawings and then mentally rotated this image until it matched the other object.

Evaluating the Kosslyn and Schwartz Theory

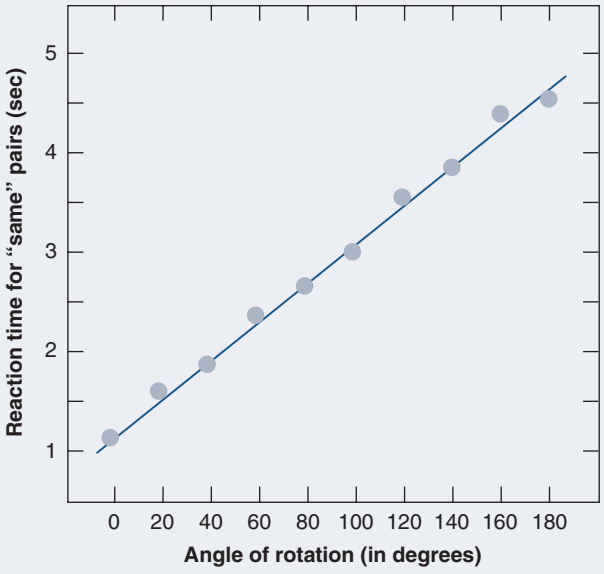
Pylyshyn (1981) provides several critiques of the theory. He says that there are too many free parameters—that is, functions or assumptions that are uncontrolled and that can

Figure 5.8 **Stimuli from an early mental rotation experiment. Participants were asked to judge whether the two patterns were rotated versions of each other.**



Source: From Shepard and Metzler (1971). Copyright © 1971 AAAS.

Figure 5.9 Response time plotted as a function of the angular difference between the two patterns suggests a mental rotation.



Source: From Shepard and Metzler (1971). Copyright © 1971 AAAS.

affect the results of any experiment designed to test the theory. For instance, Pylyshyn points out that the transformation processes that are hypothesized to act on an image are very flexible. This flexibility, rather than being a strength, leaves many questions unanswered—for example, "Under what circumstances will a given transformation process operate?" Pylyshyn would like to see a more detailed theory and a greater number of constraints put on its operation. He sees the theory as too open and as allowing for too many uncontrolled possibilities.

Pylyshyn also makes the argument that participants performing a mental rotation task may not be rotating at all. They may, in fact, be employing nonanalog or propositional means of carrying out image transformation. The data that appear to support image rotation are not reliable, according to Pylyshyn, and have been obtained because the task and the instructions led participants to re-create, as accurately as they could, the perceptual events that would have occurred if they had been observing an analog process. In other words, the observers are doing what the stimuli, task, and experimental situation seem to demand of them. This phenomenon is believed to be a potential source of error in many psychology experiments and goes by the name of **demand characteristics**.

Kosslyn counters these arguments. He makes the distinction between parameters and free parameters, and he points out that his theory contains many of the former but not many of the latter. This theory, or any general theory that explains new cases, must include many parameters. Specific theories have fewer parameters but can explain only a limited number of cases. Second, Kosslyn points out several experiments in which task demands affected the results but argues that these studies in no way preclude the use of analog imagery processing. The most compelling evidence for mental visual imagery involving analog perceptual representations, rather than logical propositional ones, comes from cognitive neuroscience. Evidence indicates that both visual perception and visual imagery seem to draw on the same underlying neural machinery. For instance, Farah (1988) showed that patients who had suffered damage to the visual cortex as the result of a stroke could no longer form visual images. Goldenberg, Podreka, Steiner, and Willmes (1989) showed that when participants are asked to answer questions that require the use of visual imagery, there is a greater increase in their vision-related brain activity, measured as increased blood flow to the visual cortex, than in participants who are given questions that do not require the use of imagery. Kosslyn, Thompson, Kim, and Alpert (1995) also showed that increases in blood flow in visual cortex were associated with visual imagery, and they even found that small mental images were focused in primary visual cortex (where receptive fields are small), whereas large mental images elicited activity in later regions of visual cortex (where receptive fields are larger).

CONCEPTS: HOW DO WE REPRESENT OUR KNOWLEDGE OF CONCEPTS?

Conceptual representation describes the mechanisms by which we store our knowledge of concepts and categories; everything from what a *dog* is, to what *respecting* means, to what *justice* is. Classical theories of conceptual representation (originally inspired by the work of Aristotle) treated concepts and categories as comprising definitions, almost like a dictionary. A concept or category would be made of a list of defining features. If a particular exemplar, such as a dog, has those defining features (e.g., four legs, a tail, it barks), then it is considered fully a member of that category. If something does not have those defining features, then it is not at all a member of that category. In the classical theory of concepts and categories, membership was all-or-none.

The most important revision to the classical theory of concepts was that defining features often do not work for categorizing certain concepts. For instance, the philosopher Wittgenstein (1953/2009) pointed out the concept of a *game* does not have a list of defining features. Although "fun to play" might be a necessary feature of games, it is not sufficient on its own because there are many nongame things that are fun to play at. Moreover, since different types of games have very different objectives, rules, and participants, creating a list of necessary and sufficient defining features of a game is essentially impossible. Instead, Wittgenstein recommended treating concepts and categories as having *family resemblances*. For example, chairs typically have four legs but some have three or five, and chairs typically have backs but some do not. A beanbag chair

and a wheelchair have almost no features in common, and yet they both have family resemblances to other chairs that allow them to be generally treated as members of the category *chai*r, more or less.

Without necessary and sufficient defining features for a concept, membership in such a category is no longer all-or-none. Eleanor Rosch (1973, 1975) conducted a series of cognitive psychology experiments that further indicated the classical theory needed substantial revision. Rosch found that people exhibit a remarkable consistency in which features they treat as indicative of a particular concept, almost as if they were accessing a common dictionary entry in their minds. However, she also found that people appear to not rely on *defining* features for forming their concepts and categories but instead use *typical* features to determine graded membership in categories (see also Rips, Shoben, & Smith, 1973). An exemplar like *robin* is a highly typical member of the category *bird*, making it a central member of the category, because it has many of the typical features of a bird, such as wings, a beak, flying, and singing. By contrast, a penguin neither flies nor sings, thus making it an atypical exemplar, and therefore not a central member of the category. Without defining features determining whether an exemplar is a member of a category or not, it appears that people treat categories as having graded membership (not all-or-none membership).

Not only can one estimate the centrality of a feature by seeing how consistently people include it in their description of a concept, one can also estimate a feature's centrality to that concept by determining what the consequences are of removing that feature. Frank Keil (1989) found that some conceptual features are more immutable (unable to change) than others. For example, a robin that doesn't chirp is still pretty much a robin, but a robin that doesn't have wings is substantially less of a robin. Sloman, Love, and Ahn (1998) found that concepts and their features can be described as adhering to a constraint satisfaction account wherein some features are immutable (central or fundamental to the category) and other features are somewhat more mutable (able to be changed while preserving category membership). They further determined that some features are dependent on other features. For example, a chair's being comfortable depends on there being a seat, a back, and arms. However, a chair's having a seat does not depend on any of the other features. Thus, the feature of having a seat is not only immutable for a chair, but it is also less dependent on other features, thus making this feature highly central to a chair. Rehder and Kim (2006) have further suggested these central features (which tend to be independent of the other features) often play a causal role in making the object into an exemplar for that category. For instance, *having a seat* can be seen as what principally causes an object to be categorized as a chair. When an object that is not chair, such as a log, has a seat carved into it, suddenly it is now an exemplar of the category *chair*. With concepts like robin and chair, the immutability, independence, and causal role of a feature turns out to be a strong predictor of how central that feature is to the concept.

In addition to there being certain dependency relations among concept features, there are also correlations among features, irrespective of their concepts. For example, most things that have two eyes also have a mouth, and most things that have fur also have whiskers. Therefore, those pairs of features tend to co-occur with each other. These correlations between features on their own may make a feature behave differently as

part of a conceptual representation. For example, imagine that the features "has a beak" and "is brown" are equally common and consistently listed for the concept *robin*, but "has a beak" is highly correlated with some of the other features of that concept (such as "has wings"). Things that have a beak also tend to have wings. Malt and Smith (1984) suggested that this property might make the feature "has a beak" more effective at activating the concept of *robin* than "is brown"—despite those two features being equally common as features of robin. McRae, de Sa, and Seidenberg (1997) tested the effect of these correlations among features in a series of reaction time experiments. With speeded tasks, they found that correlations among features reliably influenced reaction time. For example, participants performed a feature verification task, where they might see the concept "scooter" and then the feature "has wheels." The feature "has wheels" is highly correlated with several other features of scooters, such as "people ride on it" and "goes fast," so participants responded quickly in verifying that feature. However, when the same feature was presented after the concept "cannon," people were slower to respond because the feature "has wheels" is not highly correlated with other features of cannons. McRae et al. also used an artificial neural network to simulate these results, showing how conceptual representation could be carried out by connections among features and connections between those features and the concept names—without the need for dictionary-like entries.

The findings of Rosch and Sloman and McRae pointed to a revised account of conceptual representation that has managed to soften the *defining* features into *typical* features and relax the definition of category membership to be graded instead of all-or-none. However, it still treated a concept as something like a collection of features stored inside the mind waiting to be used or accessed. But what about categories that you make up "on the fly" (i.e., ad hoc categories)? Lawrence Barsalou (1983) asked participants to list items that they would remove from a burning house or what things they might sell at a garage sale. To his surprise, there was remarkable consistency across participants, similar to that found by Rosch and others for standard concepts and categories. Given that these ad hoc categories probably do not have "dictionary entries" stored inside the mind already, perhaps the remarkably consistent features listed by different people should not be taken as evidence for a concept having a preexisting stored representation. In fact, Casasanto and Lupyan (2015) have suggested that *all* concepts and categories might not be stored in the mind like dictionary entries patiently waiting to be accessed but are instead constructed "on the fly" uniquely for each context (see also Yee & Thompson-Schill, 2016). Rather than there being a central set of features for a concept that gets accessed when that concept is used, which then gets partially contextualized for a given situation, perhaps the contextualization happens first and the conceptual representation emerges from that, forming anew each time. If the context is the same across different people, then they should generate remarkably consistent descriptions of those concepts, not because they are accessing a preexisting representation but because they are building "on the fly" from the same context.

If concepts and categories are not like dictionary entries, not made of static representations stored in the mind, then what are they made of? Lawrence Barsalou (1999) suggests that they are made of **perceptual simulations**. Perhaps a concept is not an abstract representation stored in the prefrontal cortex (the reasoning and planning

region of the brain), with its features listed inside it in an *amodal* description (i.e., a logical nonsensory format). Rather, a concept may be a set of connections between neural representations in the prefrontal cortex (for its linguistic label), visual cortex (for its visible features), auditory cortex (for its audible features), motor cortex (for its action-based features), olfactory cortex (for its chemosensory features), and so on. Those sensory and motor representations are called *modal* because they are specific to that particular sensorimotor modality. If a conceptual representation is not a preexisting amodal list of features but is instead a connected collection of modal representations across many brain regions, then different contexts should generate substantially different descriptions of the same concept. For instance, think of the features that make up a *watermelon*. "Green" and "oval" may show up early in your list. However, when Wu and Barsalou (2009) asked people to list features that make up *half a watermelon*, they got very different results. Features like "red" and "seeds" showed up near the top of the list. Results like this suggest that people may not be accessing a stored amodal representation of the concept of watermelon, reporting its features, and then finally adding additional context (such as being cut in half). Instead, they may be generating a modal representation of the concept by activating sensory representations that incorporate context and central features all at the same time: an ad hoc perceptual simulation. We will see more of these perceptual simulations in Chapter 14, where cognition is seen as *embodied* (not happening solely inside the brain).

What about concepts that do not have a lot of obvious sensory features, such as abstract concepts? How would something like Barsalou's (1999) perceptual simulation framework account for concepts that are not particularly perceptual, such as *respect* or *justice?* Allan Paivio's (1990) dual coding theory has proposed that, while concrete concepts may be represented in a visual imagistic format, abstract concepts may instead be represented in a more verbal format. For abstract concepts, the representation may be less grounded in perceptual simulations and instead encoded via relationships among symbolic representations. Statistical accounts of word meaning have been proposed that can handle exactly those relationships among symbols in order to represent meanings (Landauer & Dumais, 1997; Lund & Burgess, 1996). Abstract concepts may derive their meanings not from sensory impressions that are grounded in the world but instead from statistical relationships with other already known concepts. However, Glenberg and Robertson (2000) reported evidence that at least for concrete concepts, the representation clearly must rely in part on perceptual simulations. They showed that a purely statistical account of word meaning was unable to tell the difference between sensible sentences such as "After wading barefoot in the lake, Erik used his shirt to dry his feet" and nonsensical sentences such as "After wading barefoot in the lake, Erik used his glasses to dry his feet." A major part of how you know that the second sentence doesn't make sense is because your perceptual simulation reveals that the action doesn't work properly.

To accommodate these mixed results, Max Louwerse (2011, 2018) introduced the symbol interdependency hypothesis to explain how statistical accounts and perceptual simulation accounts of concept representation can coexist. Louwerse, Hutchinson, Tillman, and Recchia (2015) examined the effect sizes of 126 perceptual simulation experiments and determined that statistical relationship effects in these tasks tend to explain much of the results when the response is speeded, whereas perceptual

simulation effects tend to explain the results when the response is slower and more deliberate. Accordingly, Anna Borghi and colleagues (2018) have suggested that, in adults, concrete concepts may be mostly encoded via links to sensorimotor impressions but partially encoded via statistical patterns in language use, and abstract concepts may be mostly encoded via statistical patterns in language use but partially encoded via links to sensorimotor impressions. For example, the abstract concept of *anger* may be mostly understood in terms of its statistical relationships with other related concepts such as irritability, aggravation, resentment, and rage. However, it is also partly understood in terms of perceptual simulations such as the conceptual metaphor of *heated fluid under pressure*—as when people say, "He blew his lid" or "She was just letting off steam" (Lakoff & Kövecses, 1987).

Debates continue today over what format of representation is used to encode knowledge about concepts. Nonetheless, there is movement in the field toward acknowledging simultaneous roles for *both* statistical symbolic accounts and perceptual simulation accounts of that representational format. We will revisit conceptual representation in Chapters 7 and 8, where we address the network approach and the evolution approach.

PROBLEM SOLVING: HOW DO WE SOLVE PROBLEMS?

Problem solving is the transformation of a given situation into a desired situation or goal (Hayes, 1989). Problems are solved not only by people but also by computers. The former type of problem solving is within the purview of the cognitive approach and is studied by experimental psychologists. Researchers in artificial intelligence and robotics study machine problem solving. We will postpone a discussion of their perspective on problem solving until later in the book.

One approach to problem solving was the Gestalt perspective, described in Chapter 3. You may recall that the Gestalt psychologists focused on insight learning, in which a solution suddenly manifests itself, perhaps as a result of some amount of subconscious processing. In this section, we focus on an approach to problem solving that has been influenced by the discipline of artificial intelligence. The approach relies on the setting of subgoals and the testing of different possible solutions in the pursuit of a solution.

Anderson (1980) lists four characteristics of problem solving:

1. *Goal directedness:* Problem-solving behavior is directed toward the attainment of a **goal**, which is the desired end point or solution to the problem. If we were attempting to solve an arithmetic problem—say, 25 + 36—our goal would be to find the sum of these two numbers.

2. *Sequence of operations:* This problem-solving behavior involves a sequence of steps. To add 25 + 36, we can first add 20 + 30, to obtain 50. This would be Step 1. We would then add 5 + 6, to get 11, in Step 2. We would then add 50 + 11, to get 61, in Step 3, the last step.

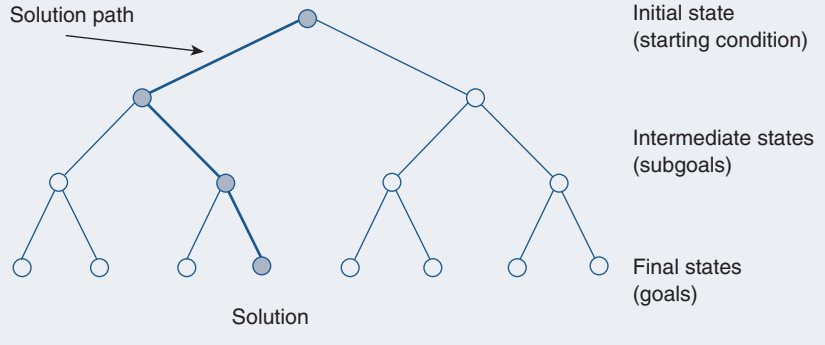
3. *Cognitive operations:* This is the application of some process to the problem-solving situation that transforms the situation. Each permissible cognitive action that is applied in this way is called an **operator**. Operators in arithmetic

include the addition of two numbers, the carrying of a remainder, or the retrieval of a number from memory.

4. *The setting of subgoals:* Each step in the problem-solving sequence produces a **subgoal**, an intermediate goal that is set along the way to an eventual solution of the problem. There is thus a hierarchical organization of goals in problem solving. The overall goal is broken down into subgoals. Each subgoal may in turn be broken down further, into additional subgoals. In our example, adding 20 + 30 yielded 50 in Step 1. This would generate the subgoal of adding 5 + 6, which produces the new subgoal of adding 50 + 11 in pursuit of the final goal, the total sum.

Problem space is a very important concept in problem solving. A **problem space** may be thought of as different situations or states that can exist in a problem. For any problem space, one can define three types of states. The initial state is what the problem is like at the outset. The intermediate states are what results after some action is applied to the problem. The goal state is a desired situation and can include the final situation or configuration of the problem after a solution has been obtained. A problem space also includes the problem solver's knowledge at each step of the problem-solving sequence, knowledge that is applied at the step being negotiated, and knowledge that is in memory and could be applied. Problem spaces are usually represented in a diagram, called a solution tree, that shows all the possible steps that can be taken in pursuit of a solution (see Figure 5.10). The solution tree contains a series of branches. Each fork that has two or more branches represents the possibility of choosing one course of action over another. A given path can ultimately lead to a dead end in a nonsolution. The correct solution to a problem is a particular path or a set of paths if there is more than one procedure that produces a correct solution.

Figure 5.10 A representation of a problem space showing the solution path.



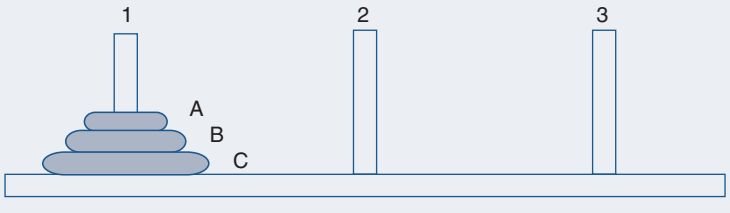
One approach to problem solving is to try all possible paths represented in the solution tree. This brute-force method is computationally intensive and time-consuming. As an example, the game of chess has been estimated to have 10^{120} states, or nodes, in its full solution tree. It is simply not possible to search through *all* of these states to locate the best path. Current computers can search several billion states per second, still far fewer than the total number of states. Chess grandmasters usually don't consider more than 100 possible states, and an average player might consider only a few dozen.

What is needed, then, is some guide to tell us which part of the space is likely to yield a correct answer. Heuristics serve this function. A **heuristic** is an informal "rule-of-thumb" method of problem solving that does not guarantee a solution, but it narrows the search to a subset of the solution tree and therefore is faster and easier to use than an exhaustive search of the entire problem space. Imagine that you have misplaced your keys somewhere in your apartment. You could engage in an exhaustive search and look everywhere for the keys: in each room, under the bed, on the shelves, behind the drawers, and so on. Alternatively, you could look first in those places where you usually put the keys down: in a desk drawer, on a tabletop near the door, or some other likely location. This heuristic would usually enable you to locate the keys more quickly and with less effort.

The most effectual heuristic in problem solving is the **means–ends analysis**. In this technique, the problem is solved via successive determinations of the difference between the existing state and the goal or subgoal state and then the identification and utilization of an operator that reduces this difference. We all use means–ends problem solving in our everyday lives. If you wanted to get groceries for the week (the goal), you would need to first get to the supermarket (a subgoal). If you were at home, you would need to drive to the supermarket (an operator). If this were accomplished, the first subgoal would, therefore, have been attained, and the difference between the state of being at home and that of needing to get to the market would be reduced. Once you were at the market, you would set up another subgoal of getting the food, which would require a second operator: obtaining a shopping cart. This process would continue until the final goal had been reached.

The Tower of Hanoi problem has been studied quite thoroughly by investigators and serves as a good example of the application of the means–ends analysis. In this problem,

Figure 5.11 The three-disk Tower of Hanoi problem.



there are three pegs, marked—from left to right—1, 2, and 3 (see Figure 5.11). Initially, there are three disks of different sizes—A, B, and C—that have been stacked on Peg 1—the largest disk at the bottom, the smallest on top. The goal is to get all three disks on Peg 3, stacked in exactly the same way. There are two strict constraints: (1) You can move only one disk at a time and only to another peg, and (2) you cannot place a larger disk on top of a smaller one. Try to solve the Tower of Hanoi problem before reading on.

Figure 5.12 shows the seven-step solution. The first goal is to free up Disk C, since it must always be at the bottom. Doing this entails setting the subgoal of getting Disk B off Disk C. However, this necessitates the setting of another subgoal, getting A off B. The first move is thus to transfer A to Peg 3, which then allows us to move B to Peg 2. The next few moves are along these lines, with the generation of new subgoals and new operators for their attainment. We cannot move C to Peg 3 just yet, because A occupies that position. To remove A, we place it on top of B on Peg 2. This frees up Peg 3 so that we can now place C there. The remaining moves are straightforward. A goes to Peg 1, B goes to Peg 3 on top of C, and A can finally be placed atop B to complete the attainment of the goal.

The General Problem Solver Model

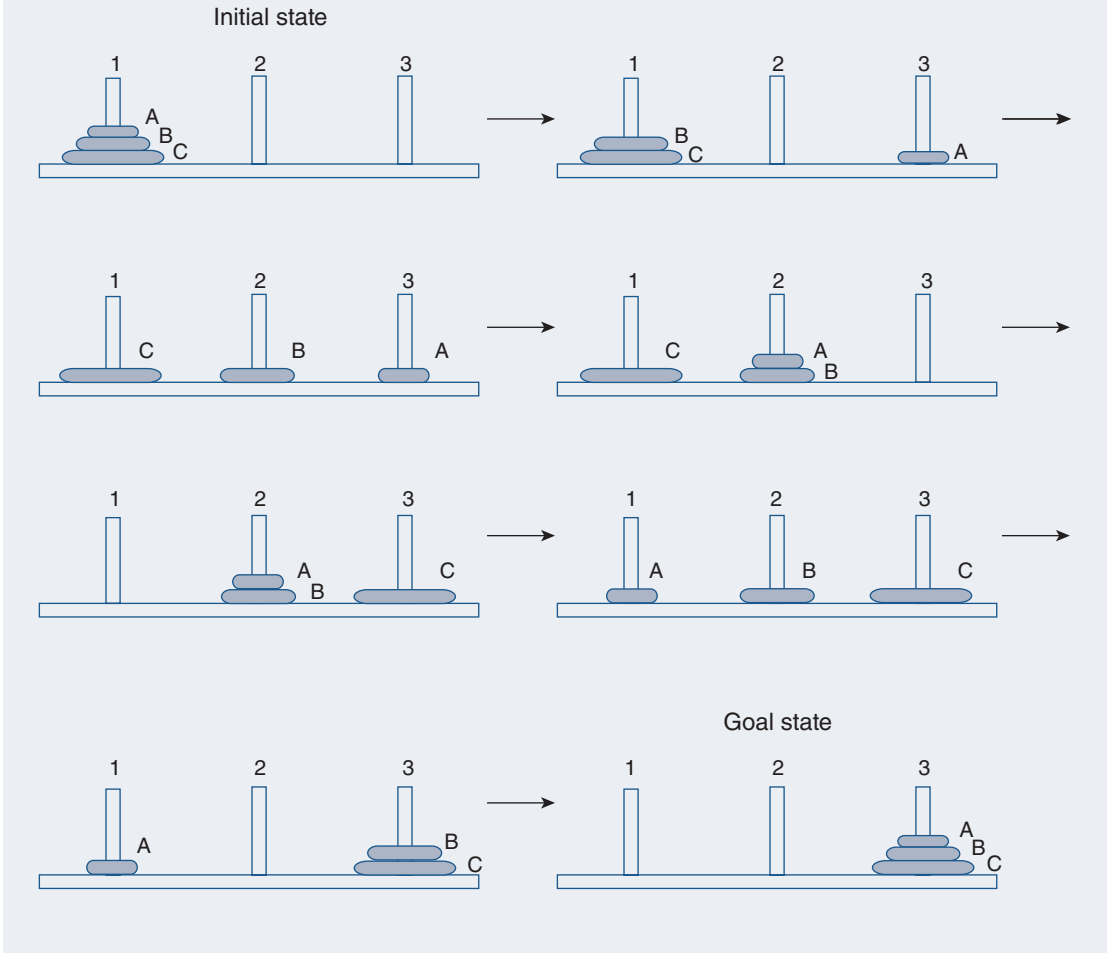
Alan Newell and Herb Simon were early pioneers in a computational approach to problem-solving research. They developed the first comprehensive computer simulation of human problem solving (Newell & Simon, 1972). Their program applied means–ends analysis to a wide range of problems, and thus, the program earned the name General Problem Solver (GPS). The GPS model serves as a useful analogy for how humans sometimes solve problems—by first representing the problem and the goal, then identifying the subgoals and operators that will produce that goal. The model suggests that humans and computers can both solve problems this way—by constructing "plans within plans" to produce a solution.

A key component of the model was its use of production rules. These rules were used to perform the operations that were required to bring about a solution. Ashcraft (2002) gives some examples of production rules for the Tower of Hanoi problem:

1. IF the destination peg is clear and the largest disk is free, THEN move the largest disk to the destination peg.
2. IF the largest disk is not free, THEN set up a subgoal that will free it.
3. IF the subgoal of freeing the largest disk has been set up and a smaller disk is on that disk, THEN move the smaller disk to the stack peg.

The GPS program has been applied to several well-known logic puzzles, including the Tower of Hanoi problem, the Missionary–Cannibal problem (described next), as well as the three coins puzzle, the water jug task, and more. The GPS program was able to solve each of these. In addition, Newell, Simon, and others have tested the model. In one study, they had participants talk out loud as they attempted to solve abstract logic

Figure 5.12 The seven-step solution to the three-disk Tower of Hanoi problem.



problems (Newell & Simon, 1972). They found that there was a significant correspondence between the steps the participants reported using in their verbal protocols and the steps employed by GPS.

Evaluating the General Problem Solver Model

However, GPS has its limitations. In some cases, it diverges from the strategies human beings use to solve problems. With some problem solving, human beings are apt to use another heuristic, called the hill-climbing strategy. It involves taking actions that always bring one closer to the goal. To illustrate, if you were climbing a mountain and your goal

was to get to the top, you would always choose a path that led upward. This strategy can sometimes fail, however—in this example, in the case of a path that goes downward and then back upward toward the summit. In the Missionary–Cannibal problem, there are three missionaries and three cannibals on one side of a river. The goal is to move all the missionaries to the opposite side by a boat, two at a time. If there are more cannibals than missionaries on one side, they get eaten. To solve the problem correctly, there is a necessary step where one must carry people who have been transported to the far side of a river back to the original side of the river. Because this step leads away from the final goal, which is to move everyone to the far side, persons attempting to solve the problem do not take the step and, therefore, fail to solve the problem. The GPS program doesn't have difficulty with this particular kind of pitfall but does sometimes fail to obtain a solution because it applies the means–ends heuristic too rigidly (Greeno, 1974).

INTERDISCIPLINARY CROSSROADS: ARTIFICIAL INTELLIGENCE, PROBLEM SOLVING, AND THE SOAR MODEL

SOAR—which historically stood for "state, operator, and result"—is a problem-solving model, but it is also a universal cognitive architecture, a system designed to account for a wide range of cognitive phenomena (Newell, 1991). **Cognitive architectures** specify the structure and function of many different cognitive systems and how their structure and function interact. Broadbent's (1958) early model of attention and Atkinson and Shiffrin's (1971) modal model of memory are examples of cognitive architectures. These models can be applied within a specific domain—such as attention, memory, or problem solving—but they go beyond this, as they attempt to describe the basic principles of cognitive functioning.

As we will see, SOAR incorporates some of the elements found in GPS but is more general. Let's discuss its main features. SOAR represents all its tasks as problem spaces. Any problem is broken down into an initial state (the problem's initial starting point), a desired state (the goal or solution), and the current state (what the problem is like at any given moment). It can apply any one of a set of operators to the current state, thereby altering it to create a new state that is closer to that of the desired goal. SOAR was tested on block-world problems. A block-world problem consists of a set of blocks that has one configuration—a starting configuration—that must be rearranged so as to yield a desired end-state configuration (see Figure 5.13).

In SOAR, knowledge stored in long-term memory is used to select operators and guide movement through the problem space. The memory consists entirely of production rules that specify an action that is to be taken if a set of preconditions is met. Objects, such as blocks in the block-world problem or disks in the Tower of Hanoi problem, are represented by attributes and values. These are simple descriptive characteristics that can be used to define objects exclusively in the problem scenario.

(Continued)

(Continued)

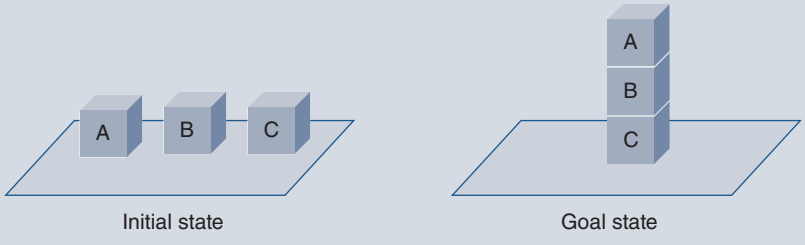
SOAR goes through a decision cycle, where it accumulates evidence that has to do with the problem. Once this has taken place, a decision is executed. Examples of decisions that can be made are which problem space is to be used, which state is to be used within a given space, and which operator is needed to enable progression to the next state. One of the model's most interesting aspects is its capacity to develop preferences as it is accumulating evidence. These preferences are for actions that ought to be taken, given their likelihood of bringing about the goal. Based on these preferences, a given operator may be accepted, rejected, or considered as better, indifferent, or worse.

SOAR, like other computer programs, can follow a preprogrammed set of instructions for solving a problem. But, unlike these programs, it is also capable of generating new ways of approaching the problem. SOAR creates novel subgoals when it comes up against an impasse—a situation in which there is no option better than the rest. The setting of the new subgoal is based on knowledge stored in memory. This ability allows SOAR to adapt itself to the demands of a novel situation flexibly and dynamically.

Figure 5.14 gives an example of how SOAR would solve a simple block-world problem. The initial state has Blocks A, B, and C arranged in a row on a table. The goal state is a reconfiguration of the blocks such that Block A is on top of Block B and Block B is on top of Block C. At the initial starting point, there are three permissible operators. A can go on top of C (A → C), B can go on top of C (B → C), or A can go on top of B (A → B). SOAR evaluates these three options by trying them out in an evaluation space. The A → C option produces a configuration that leads to an unacceptable solution, with B on top of A and A on top of C. The A → B option also produces an unacceptable state, with C on top of A and A on top of B. SOAR then selects and implements the only remaining operator to produce a successful solution.

A final comment on the SOAR model concerns chunking, which in this context is comparable to the process of grouping items together in memory, discussed earlier. Chunking in this context corresponds to learning from experience. If SOAR has tested possible solutions and found a particular solution to a problem in the past, it stores this solution in a general form in its memory. When SOAR encounters a similar problem at a later date, it can then apply operators it has applied previously to reach a similar solution. SOAR, in effect, recognizes the problem as a combination of conditions that trigger the appropriate action. This grouping or chunking of the conditions into a single precondition for applying an operator gives the process its name.

Figure 5.13 A typical block-world problem.



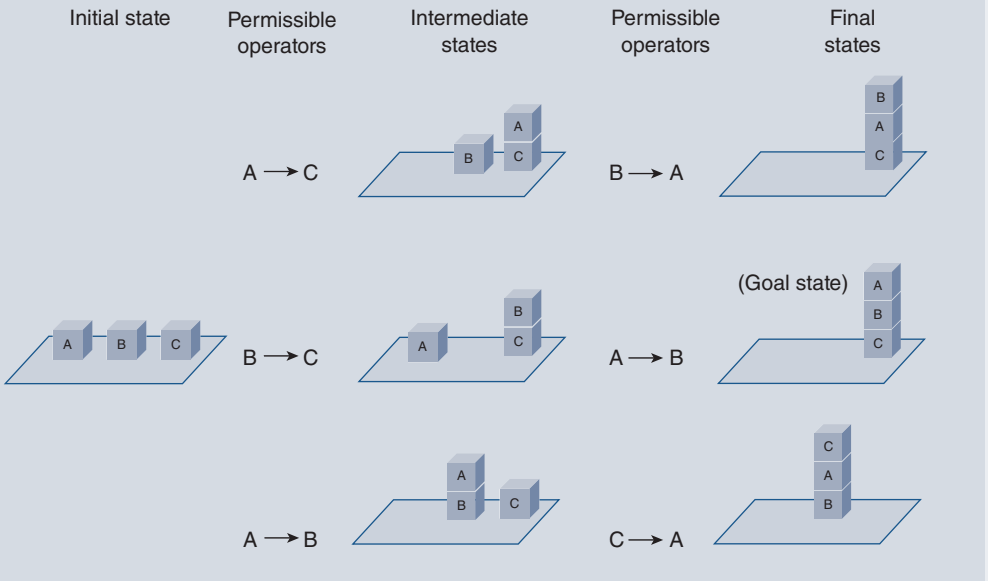
Evaluating the SOAR Model

The SOAR program represents an ambitious attempt to develop a cognitive architecture that does more than solve problems. It was designed as a model for delving into the general principles that underlie learning, memory, problem solving, and the interactions among these systems. It has a good track record, having been applied to and having solved a variety of problems.

What's more, the incorporation of the chunking feature parallels the way human experts diagnose and solve problems. Chase and Simon (1973) presented pictures of a chessboard whose pieces occupied random positions and pictures of a game in progress to two groups of participants—expert chess players and novice chess players. There was no difference between the recall of the novice players and that of the expert players when it came to remembering the pictures of the randomly placed chess pieces, but experts were better than novices when it came to remembering the game-in-progress pictures. This suggests that the experts were chunking information; they were organizing portions of the information into meaningful wholes. These wholes then guided the selection of the appropriate operator to be applied to the problem. For the expert chess players, seeing a particular configuration of pieces apparently triggered a memory of how that situation was dealt with in the past. Auto mechanics and physicians also chunk when they diagnose the cause of a mechanical breakdown or biological illness from a set of symptoms.

A critique of the model centers on the assumption that only production rules make up memory. Other architectures posit separate memory types with separate structures that code for declarative and procedural knowledge. The SOAR view of a single unitary type of memory simplifies the model but runs contrary to evidence that supports the existence of distinct memory systems.

Figure 5.14 How SOAR might solve a block-world problem.



OVERALL EVALUATION OF THE COGNITIVE APPROACH

The strengths of the cognitive approach are many. Unlike the disciplines that preceded it, cognitive psychology is governed by a set of common assumptions. These assumptions are that the mind represents and operates on information and that specific models of how the mind does these things can be formulated. Another advantage of the cognitive approach is that—rather than viewing the mind in a general, undifferentiated way, as some of the adherents of other approaches did—it compartmentalizes the mind into distinct processes. This carving up of the mind into modules makes research easier because it simplifies and delimits the range of a phenomenon under investigation. Although not all cognitive processes may actually be compartmentalized and modular in their function, this compartmentalization for pragmatic purposes can allow researchers to make more advances in their study of memory, attention, conceptual representation, or problem solving.

Another advantage of the cognitive approach, as discussed at the end of the previous chapter, is its use of model building. Models provide a concise way of formulating how a given mental process might operate. They are more precise than theories and can be tested using experimental data. The implementation of a theory as a model and the subsequent modification of the model on the basis of empirical findings is a fruitful method for uncovering more about a cognitive process.

Of course, like any other discipline, the cognitive approach has its limitations. It supports the traditional view of information processing, in which representations are symbolic and transformations are operations that are performed on the symbols. This needs to be reconciled with the network view—that information is coded in a distributed way and that transformations consist of activations of nodes in a network. Model builders in cognitive psychology also need to consider findings from neuroscience. Any model maker that attempts to provide an information-processing explanation of a brain function should map the different components of the relevant model onto the known anatomical and physiological evidence. This will help delineate and develop the model so that it reflects its biological implementation more accurately.

SUMMING UP: A REVIEW OF CHAPTER 5

1. Memory is the capacity to retain information over time. Without memory, a human, animal, or machine could not learn. There are three major types of memory: sensory, working, and long term.
2. Sensory memory is where information from the senses is held very briefly so that it may be processed. There are as many different forms of sensory memory as there are sense modalities. However, we focus on the two most studied forms: iconic and echoic.
3. Iconic memory holds a brief visual representation of whatever it is you have just seen, whereas echoic memory holds a brief

auditory representation of whatever you have just heard. Researchers test iconic memory by presenting a letter array and then cueing participants to a portion of the array to report what they see. The results show that capacity is unlimited; it can hold a representation of an entire scene but only for less than a second.

4. Working memory holds the contents of one's thoughts in immediate awareness to allow for reasoning and problem solving. It is capable of retaining only 7 ± 2 items on average for less than a minute without rehearsal. There are auditory, visual, and semantic codes.
5. There are several different types of long-term memory. Implicit memories hold procedural knowledge, which is a skill or action. This information is subconscious and is demonstrated by doing. Explicit memory is for declarative knowledge, which deals with facts (semantic) or events (episodic). This is demonstrated verbally and is a part of conscious awareness.
6. The earliest process model of memory is the modal model. It gives a very broad overview of the memory types and of the information flow between them. In particular, it outlines encoding, as the process by which information gets into the system; storage, for the preservation of information; and retrieval, whereby data may be called out of long-term storage for immediate use.
7. Baddeley's working memory model shows the sorts of interactions that take place between the components of working memory, which are the visuospatial sketchpad, the articulatory loop, and the executive control system.
8. According to Kosslyn, visual images preserve the same spatial characteristics as the real-world objects or scenes they depict. This means that it would take longer to scan across a visual image of a large object than to scan across a visual image of a comparatively short object.
9. Rotation is another transformation that could be applied to a visual image. Mental rotation studies show that it takes longer to match a figure against another the more one is rotated relative to the other.
10. Problem solving involves taking a given situation and transforming it into a desired situation or goal. Many of us will resort to using heuristics to solve problems. A heuristic is a "quick-and-dirty" solution that has worked in the past but may not guarantee a correct solution. Two early computer problem-solving models are GPS and SOAR.

SUGGESTED READINGS

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Baron, J. (2007). *Thinking and deciding.* Cambridge, England: Cambridge University Press.

Kosslyn, S. M., Thompson, W. L., & Ganis, G. (2006). *The case for mental imagery*. Oxford, England: Oxford University Press.

Newell, A. (1994). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.

Radvansky, G. (2005). *Human memory.* Boston, MA: Allyn & Bacon.

CHAPTER SIX

THE NEUROSCIENCE APPROACH

Mind as Brain

THE NEUROSCIENCE PERSPECTIVE

Neuroscience is the study of nervous system anatomy and physiology. It is concerned with both the structure as well as the function of the nervous system in humans and other animals. As such, neuroscience provides a body of knowledge that serves as a foundation for understanding how cognitive operations are carried out. It gives us a hardware or implementational level of description on which we can base an algorithmic and computational description. Neuroscience is invaluable in this regard because it constrains the sorts of models of brain function that can be formulated in the other approaches.

There has been a fairly recent trend in neuroscience toward the integration of biology with cognition. Out of this union, a new discipline has emerged, called **cognitive neuroscience**. The goal of this discipline is to explicate the structures and physiological processes that underlie specific cognitive functions. We adopt a cognitive neuroscience perspective in this chapter and focus on the cognitive operations that were surveyed in the two chapters on the cognitive approach (Chapters 4 and 5). These cognitive operations are visual object recognition, attention, memory, concepts, and problem solving. This neuroscience perspective builds and expands on the information-processing models for these operations that were described previously. For neuroscientific accounts of visual imagery and language, please see the relevant sections in Chapters 5 and 9, respectively.

A fair amount of background knowledge is needed to interpret the findings of neuroscience studies. We will, therefore, provide general information on methodology, anatomy, physiology, and terminology prior to a discussion of the specific topic areas. Our present discussion begins with the tools used by neuroscientists and is followed by a discussion of the anatomy and function of neurons and of the overall brain. We will then address the neuroscience of cognitive functions and models of their operation.

Learning Objectives

After reading this chapter, you will be able to:

1. Contrast electroencephalography (EEG) with functional magnetic resonance imaging (fMRI).
2. Compare transcranial magnetic stimulation (TMS) with electrical stimulation techniques.
3. Describe the anatomy of a typical neuron.
4. Identify major anatomical brain regions.
5. Distinguish the dorsal and ventral visual pathways in the brain.
6. Define *distributed coding* among neurons in cortex.
7. Summarize Hebbian learning among neurons and long-term potentiation.
8. Describe the role of the hippocampus in memory.

METHODOLOGY IN NEUROSCIENCE

There are many procedures for collecting data in neuroscience. In this section, we group these procedures into three main categories. The first category of procedures involves the investigation of brain damage. In studies of humans, researchers—using the case study method—investigate the behavioral deficits of patients who have suffered damage to some part of the brain as a result of an accident. In animal lesion studies, this damage is deliberately induced. The second category of procedures involves the recording of brain activity in healthy subjects. The electrical action of individual neurons or small groups of neurons can be measured with cellular recording techniques that use an implanted electrode. Larger patterns of brain activity can be measured through the use of surface electrodes that have been implanted on the scalp or through the use of more sophisticated brain imaging devices. The third category involves direct stimulation of the brain itself.

Techniques for the Study of Brain Damage

It is, of course, unethical to damage a person's brain deliberately to look at the effects of brain damage. For this reason, researchers examine brain damage and its effects in people for whom the damage has come about as a result of an accident—what is called the **case study** method. There are different types of brain damage. In a patient who has had a stroke, a blood vessel servicing a part of the brain bursts, depriving blood supply to that part of the brain. Direct mechanical damage to the brain accompanies head trauma. This type of damage can be incurred in an automobile accident or a fall. Brain damage is also an end result of exposure to carbon monoxide gas. Usually in these cases, localized damage occurs and follow-up evaluations of affected persons can reveal exactly which areas were damaged. The resulting behavioral and cognitive deficits can then be mapped onto the damaged areas. If, for example, the front part of the brain has suffered damage and the patient has subsequently had trouble planning and making decisions, we can conclude that this part of the brain is implicated in the cognitive operations of planning and making decisions.

A related method of study involves the precise destruction of a specific brain area in an animal and the examination of the resulting behavioral deficits. A brain lesion is a wound in or injury to brain tissue. A study that uses this method is referred to as a **lesion study**. Lesions can be generated in several ways. Brain tissue can be drawn off by suction, destroyed by passing a high-frequency radio current through an electrode (or by introducing gamma rays, ultrasound waves, or certain acids), or temporarily frozen by pumping a coolant through an implanted probe. With this last procedure, the tissue is preserved, and normal function returns as the affected area warms up. The motivation for using these procedures is the same as that for using the case study, and therefore this type of study suffers from the same flaws that we outline below.

Evaluating Techniques for the Study of Brain Damage

There is logic behind case studies and lesion studies. If brain area *X* is damaged and a deficit is subsequently observed in behavior *Y*, researchers infer that area *X* plays some

role in the control of behavior *Y*. This seems straightforward enough but is actually fraught with difficulties. Some have likened this kind of conjecture to removing part of a car engine to see how its removal affects the car's function. If we removed the carburetor or fuel injector, the engine wouldn't run because the spark plugs wouldn't receive any gasoline. However, the same would be true if we removed the gas tank. The point is that in a car, many of the systems are interdependent. Removing any of a number of different parts can produce the same symptom. Areas of the brain are also interdependent. The effects of damage to one area could have a variety of functional interpretations. That area could be entirely responsible for a cognitive ability or part of a collection of areas that mediate that ability. Alternatively, the area could be a center where information is processed or simply a region through which neural pathways pass as they connect two other brain regions.

Brain Recording Methods

In **single-cell recording**, a very fine microelectrode is inserted into either a single neuron or the extracellular fluid adjacent to it. Changes in that cell's electrical conductivity or rate of firing can then be measured. In **multiunit recording**, a larger electrode is used to measure the collective electrical activity of a group of neurons. With multiple such electrodes in close proximity, the activity of several individual neurons can be resolved. In a classic study, Hubel and Wiesel (1962) recorded activity in single cells of a cat's visual cortex. They discovered that some cells responded to lines of a specific orientation in the cat's visual field, implying that line orientation constitutes a basic visual feature of pattern recognition. Although cell recording can give us very detailed information about what one neuron or a relatively small number of neurons are doing, it fails to yield the "big picture." That is, it does not inform us about more global brain activity.

If one wants to investigate what the brain as a whole is up to, he or she can choose from a variety of other tools. The **electroencephalogram (EEG)** is a recording of the brain's gross electrical action. When large groups of neurons generate synchronized firing patterns, they produce electric fields that extend out beyond the skull. This noninvasive procedure uses flat surface electrodes that are attached to the scalp of a human or other animal. The output of an EEG is in the form of electric wave patterns at a wide range of frequencies. This method has been used extensively in the measurement of brain activity during sleep. From research that has used this method, we have learned that the brain passes through several distinct sleep stages, each having its own frequency of wave activity.

The EEG can also be used to measure a subject's brain activity in response to his or her experience of a particular stimulus event. The resulting waves that are recorded are called **event-related potentials (ERPs)**. For instance, a researcher could sound a tone to a participant in a study and then record the subsequently occurring brain activity. This has been used in many studies to yield insights into the brain mechanisms that underlie perception and attention. A strength of EEG (and ERPs) is that it provides high temporal resolution, as it typically samples the electrical fields at around 1,000 Hz (a thousand data points per second). Therefore, it can detect fine-grain changes in cognitive processes at the time scale of milliseconds. However, a weakness of EEG (and ERPs) is that

its spatial resolution is relatively coarse. It shows us activity only in relatively large brain areas, such as the different cortical lobes.

More recent technologies have produced new **brain imaging** techniques that allow us to "peer inside" the head with greater spatial resolution than ever before. These techniques allow imagery of both brain structure and function. They allow us to see not only the static three-dimensional organization of brain areas but also the dynamic activities of these areas as they unfold in time. However, some of them do not have the high temporal resolution that EEG has.

Positron Emission Tomography

Positron emission tomography (PET) scan measures blood flow in the brain while a participant is carrying out a cognitive task. This is accomplished through the use of radioactive isotopes or tracers attached to carrier substances such as glucose or oxygen molecules. The molecules are injected into the participant's bloodstream, wherein they make their way to the brain. Brain areas that are more active during the execution of a task will show greater regional cerebral blood flow (rCBF) and, correspondingly, a greater concentration of the tracer molecules bearing the isotopes. The activity in these areas is measured using a detection device that counts the positron particles that are emitted by the isotopes. PET scans have been useful for mapping those parts of the brain that are involved in specific cognitive operations, such as visual imagery, language, and memory.

PET scans have a fairly good spatial resolution, being able to monitor locational changes in brain activity to within just a few millimeters. PET scans lack some of the temporal resolution of other brain-mapping techniques, however. They cannot show rapid changes in brain activity—those that occur over time periods that are on the order of milliseconds—like EEG can. The radioactive isotopes that are used are expensive, and their radioactivity is short-lived. There is also a small risk associated with the frequent introduction of radioactive compounds into the body.

Functional Magnetic Resonance Imaging

In most **functional magnetic resonance imaging (fMRI)** experiments, a participant is placed inside a tube that contains a powerful magnet. Protons, which are subatomic particles present everywhere in the body, align themselves in the magnetic field in the same way iron filings organize themselves around a small magnet. A radio wave pulse is then applied to the brain or other part of the body undergoing the scan. The radio signals are bounced back and picked up by a detector unit. The reflected signals are then converted to video images showing dynamic changes in brain activity over time.

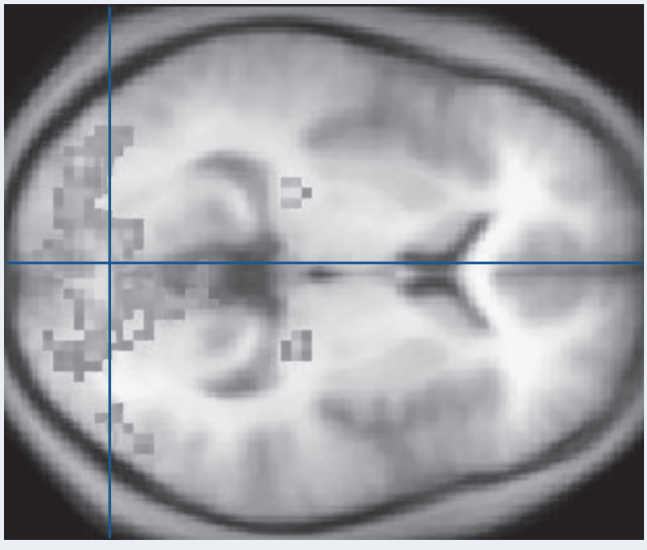
Like PET scans, fMRI scans detect alterations in local blood flow and oxygen level. Brain areas that show increases in these measures are those that have been activated during specific cognitive operations and are therefore drawing more blood to feed their metabolic processes. So fMRI scans, like PET scans, are used to map out these active brain areas. fMRI scans provide better spatial resolution than PET scans, without any of the risks associated with the injection of radioactive isotopes. Many cognitive neuroscientists use fMRI for investigation of cognitive brain functions. However, like PET, fMRI must average its imaging data over the course of a couple seconds or more and therefore

is unable to detect rapid changes in brain activity on the time scale of milliseconds the way EEG can. Figure 6.1 depicts an MRI image showing a sagittal section of the brain.

Magnetoencephalography

The electric fields that active neurons generate also have magnetic properties. In a technique called **magnetoencephalography (MEG)**, these magnetic fields are measured to determine patterns of brain activity (Schwartz, 1999). While EEG has poor spatial resolution but high temporal resolution, and fMRI has high spatial resolution but poor temporal resolution, MEG has both good spatial and temporal resolution. It can resolve neural activity in a region as small as a millimeter at 1,000 times per second. A participant first dons a helmet filled with an array of supercooled magnetic detectors. The individual then performs an experimental task while his or her magnetic activity is recorded. The resultant data can then be converted into a three-dimensional video depicting the precise locations and patterns of activation.

Figure 6.1 Results of an fMRI brain scan showing increased blood flow in visual cortex (and lateral geniculate nucleus) during viewing of a visual stimulus. The image is made by subtracting the blood flow levels of a control condition (viewing a blank screen) from the blood flow levels while viewing a moving visual stimulus. Square pixels show greater blood flow in those brain regions during the visual task, and non-pixelated regions show no difference in blood flow for the two tasks.



Source: Wikipedia, OpenStax Textbook Anatomy and Physiology. CC BY 4.0: https://creativecommons.org/licenses/by/4.0/.

Knife-Edge Scanning Microscope

A recent development in brain imaging technology is the **knife-edge scanning microscope (KESM)**. This device consists of a diamond-tipped knife that slices through tissue. A white light source illuminates the strip of tissue as it comes off the blade, reflecting an image to a camera. The subsequent video is then processed by a computer system to provide a three-dimensional reconstruction accurate down to the cellular level. The KESM is capable of digitizing the entire volume of a mouse brain (which is about a cubic centimeter) at a 300-nanometer resolution within 100 hours (McCormick & Mayerich, 2004). If machines like this were used to scan the human brain, we could record the location and connectivity of each neuron, providing a static but invaluable data set from which to extrapolate brain function.

Brain Stimulation Techniques

Electrode Stimulation

So far, we have examined procedures that measure brain damage or normal brain function. There is a third procedure that is neither as intrusive nor as passive. This method involves the actual activation of a specific brain area via electrical or magnetic stimulation. In **electrical stimulation**, an electrical current is passed through a bipolar electrode, which causes the neurons of a localized area of brain tissue to become active. The resulting behavioral effects are usually the opposite of those observed in the brain lesion technique. Lesions typically result in negative symptoms—the disruption of some cognitive processes that were normally influenced by the now-damaged region. Electrical stimulation, on the other hand, encourages neuronal firing, which can result in positive symptoms—the facilitation of those cognitive processes. A difficulty with this procedure is that stimulation might induce supranormal activity, or an overactivation of the region and the areas associated with it. This would produce behaviors not associated with normal functioning. However, prolonged electrical stimulation can also induce the equivalent of a temporary lesion. A brain area in which the neurons have been stimulated constantly for several seconds will often undergo a refractory period afterward, during which the neurons are too fatigued to respond to the input signals that normally activate them. For several seconds, that brain region may not be able to perform its functions. An interesting example of electrical stimulation comes from research in Sharon Thompson-Schill's laboratory, where she found that direct current stimulation over left prefrontal cortex (an area associated with cognitive control) impaired people's ability to categorize objects by their color or shape (Lupyan, Mirman, Hamilton, & Thompson-Schill, 2012). This reduced ability to exert cognitive control over the categorization task, when left prefrontal cortex was temporarily deactivated, turned out to produce *improved* performance in a creative task of generating uncommon uses for everyday objects (Chrysikou et al., 2013).

Transcranial Magnetic Stimulation

In **transcranial magnetic stimulation (TMS)**, a device with a wire coil is placed over a person's head. An electrical current is passed through the coil, inducing a magnetic field. This field directly causes neurons to fire. If we place the field over different parts of the

cortex, we can witness the functionality of these areas. Activation of the motor cortex causes movement of different body parts, such as the fingers (Pascual-Leone et al., 1998). If it is placed over the primary visual cortex, vision is temporarily suppressed (Walsh & Pascual-Leone, 2005). With a single brief pulse of the magnetic field, TMS can induce activation of a brain region that will facilitate that brain region's processing of incoming signals. However, with multiple repeated pulses, TMS will instead induce a refractory period that functions as a "temporary lesion" during which that brain region will be inhibited in its processing of incoming signals for several seconds. If TMS is applied over and over again for a lengthy period of time, it can be used to treat a variety of disorders, including depression (Schutter, 2008). This technique is called repetitive TMS, or rTMS.

Optogenetics

Optogenetics is an exciting new technology that allows researchers to control neurons using light (Deisseroth, 2010). It involves the use of genes that code for a category of proteins called opsins. Opsins embedded in a cell membrane regulate the flow of electrical charge across it. This in turn affects the way in which neurons signal, whether or how they transmit messages to other neurons around them. When opsin genes are inserted into neurons, the proteins are manufactured by the cell and incorporated into the membrane. However, these proteins are now light sensitive; they can be activated by particular wavelengths of light, such as blue or green.

This light sensitivity allows researchers incredibly accurate control over individual neuron behavior. They can be turned on or off, for instance, simply by shining pulses of light onto them. Moreover, this can be done at very rapid speeds, on the order of milliseconds. For example, "fast" channel rhodopsins have allowed the generation of action potentials at more than 200 times per second. Investigators have even been able to control different groups of cells at the same time, one group being stimulated by blue light and the other by yellow light.

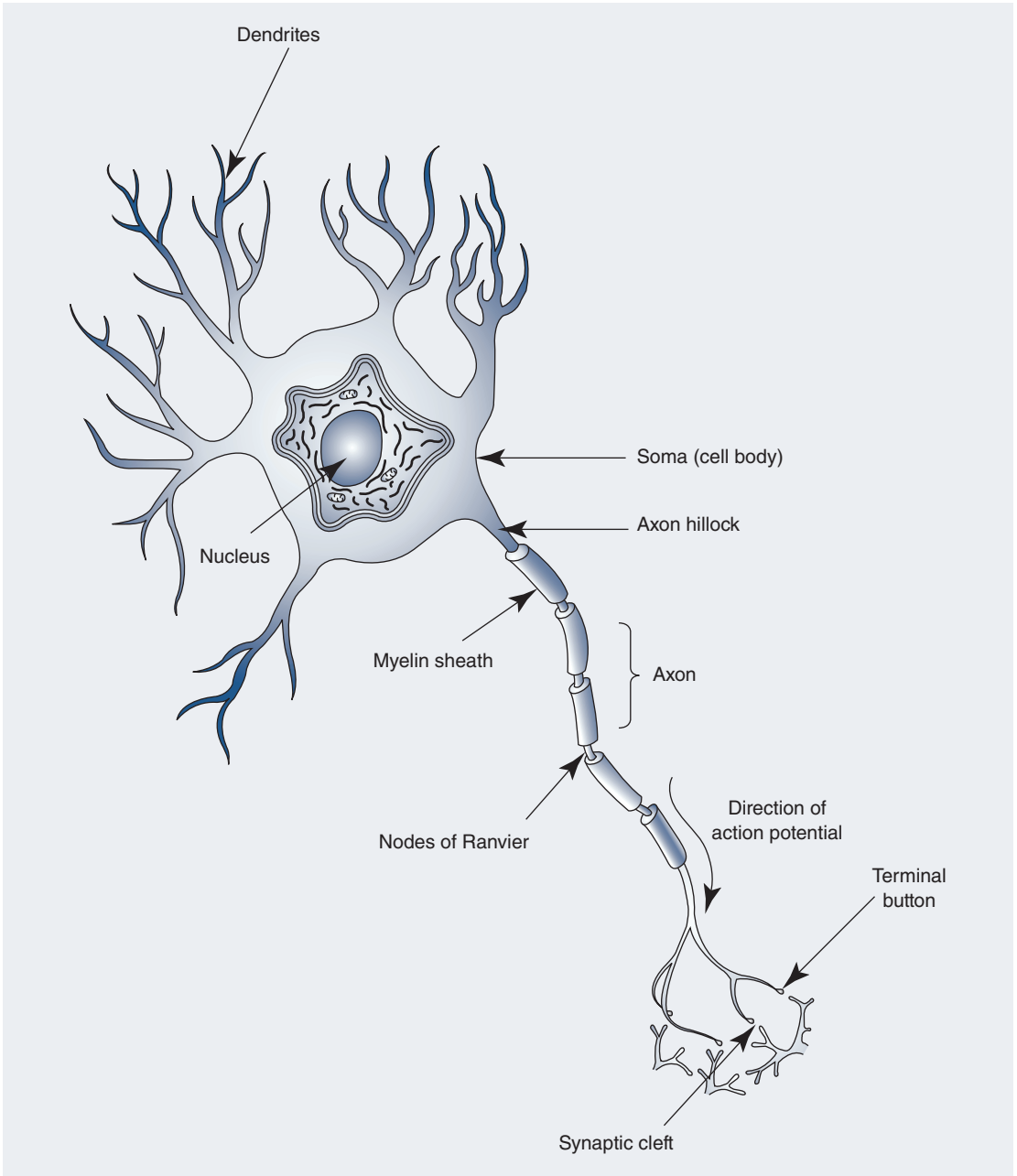
The potential for the use of this technology is enormous. Recently developed fiber-optic technology allows light to be delivered anywhere inside an animal's brain, and in conjunction with new optical recording techniques, neural tissue can be both stimulated and measured simultaneously without interference. This has been done even in moving animals and will allow us to infer with much greater precision the types of computations neural populations are performing. Rather than just triggering activity with a sensory stimulus, we can now modify neuronal activity as it is happening. Other benefits include better understanding of and promising treatments for Parkinson's disorder, schizophrenia, and autism.

THE SMALL PICTURE: NEURON ANATOMY AND PHYSIOLOGY

Neurons are the microscopic basis of the brain. They are the individual functional units that perform computations. The purpose of a neuron is to conduct a message in the form of an electrical impulse. A neuron can be thought of as a link in a complex chain because it receives incoming messages from thousands of other neurons and then sums those

excitatory (positive) and inhibitory (negative) signals to determine whether it sends an outgoing message of its own. Figure 6.2 depicts the major structures of a typical neuron. Messages are received by the feathery projections known as **dendrites**. Dendrites form

Figure 6.2 Anatomy of a neuron.

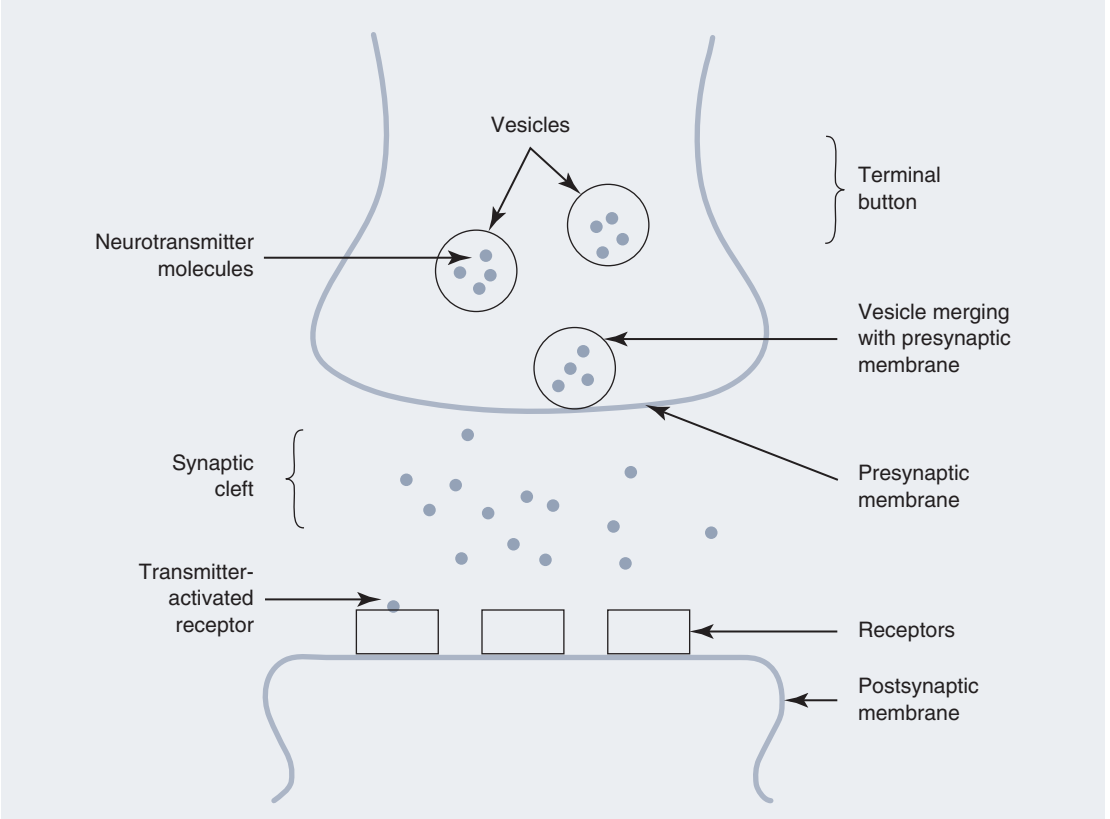


an extensive branching "tree," which connects the neuron to many other neurons. Any incoming signals picked up by the dendrites are then passed along to the cell body.

Received signals from other cells converge at the axon hillock, where the cell body meets the axon. Here, the neuron sums all the excitatory and inhibitory inputs it receives. The cell fires if the sum of these inputs exceeds the cell's **threshold of excitation**. This process represents a substantial change in the neuron's electrical state. If the threshold is met, an electrical signal called an **action potential** is initiated. The action potential then propagates down the **axon**, a long tubular structure that projects outward from the cell body. The axon, which can extend for some distance, ends in a number of **terminal buttons** that meet the dendrites of many other neurons.

The terminal buttons do not have direct physical contact with the dendrites of the other neurons. Instead, there is a gap between the two cells known as the **synaptic cleft** (see Figure 6.3). How, then, is a message passed from one cell to the next? The answer lies in molecules known as **neurotransmitters**. The job of these neurotransmitters is to complete the transmission of the signal across the synapse. When the electrochemical action potential arrives at the terminal button, it triggers the release of

Figure 6.3 Cross section of a synapse.



neurotransmitter molecules into the synaptic cleft. The transmitters diffuse across the synaptic cleft and attach to receptor molecules located on the dendritic surface of the next cell. A neurotransmitter molecule fits into a cavity in the receptor molecule and activates it, in much the same way that a key fits into and opens a lock. These activated **receptors** then contribute to the formation of a new signal in the receiving cell that propagates down the dendrite toward the cell body. There is a wide variety of neurotransmitters in the human brain. Table 6.1 gives a summary overview of some of the major neurotransmitters, what functions they subsume, and their associated disorders.

Table 6.1 Some of the major neurotransmitters, their functions, and associated disorders.

Neurotransmitter	Function	Disorder
Acetylcholine	Used to control muscles; also involved in learning	Alzheimer's disease
Dopamine	Brain reward system, control of movement	Schizophrenia, Parkinson's disease
Endorphins	Reduce pain, enhance reinforcement	Chronic pain
Gamma-aminobutyric acid (GABA)	Primary inhibitory transmitter	Anxiety, epilepsy
Glutamate	Primary excitatory transmitter, memory formation	Depression, schizophrenia
Norepinephrine	Arousal and attentiveness	Stress
Serotonin	Regulates mood and sleep	Depression, obsessive–compulsive disorder

THE BIG PICTURE: BRAIN ANATOMY

We now step back a bit and look at the brain from a broader perspective. The brain is a complex structure, with many areas and parts forming networks that are associated with particular functions. Rather than catalog all of them, we will describe only the ones relevant to important cognitive processes. For those interested in delving deeper, a number of more comprehensive texts on this topic are listed in the "Suggested Readings" section.

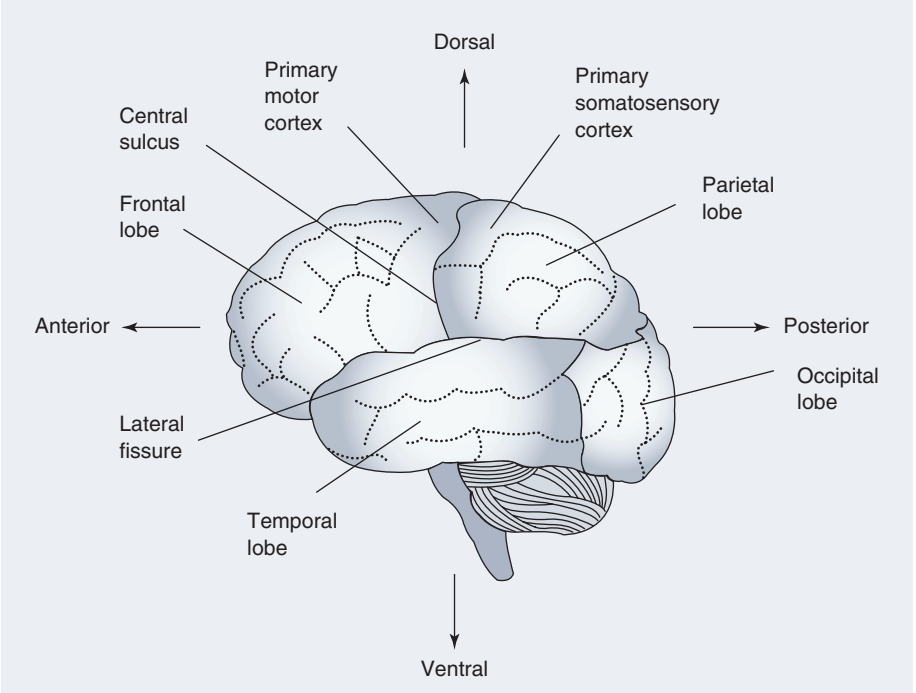
Directions in the Nervous System

Our discussion of brain anatomy must begin with a discussion of the terms used to imply direction in the nervous system. Anatomists use a special set of Latin terms when referring to these directions. With respect to the human brain, **dorsal** means "toward the top" and **ventral** means "toward the bottom." **Anterior** is used to signify regions that are toward the front of the brain, and **posterior** signifies those that are toward the back. Regions that are located toward the interior of the brain are **medial**, whereas those near the outer surface are **lateral** regions. Figure 6.4 illustrates the application of some of these terms to a view of the cortex.

The Cortex

The cortex is the most recent part of mammalian brains to evolve and is responsible for a number of higher-order cognitive abilities. The cortex is naturally divided into

Figure 6.4 Major lobes of the left hemisphere of the human cortex. Some common anatomical directions are also indicated.



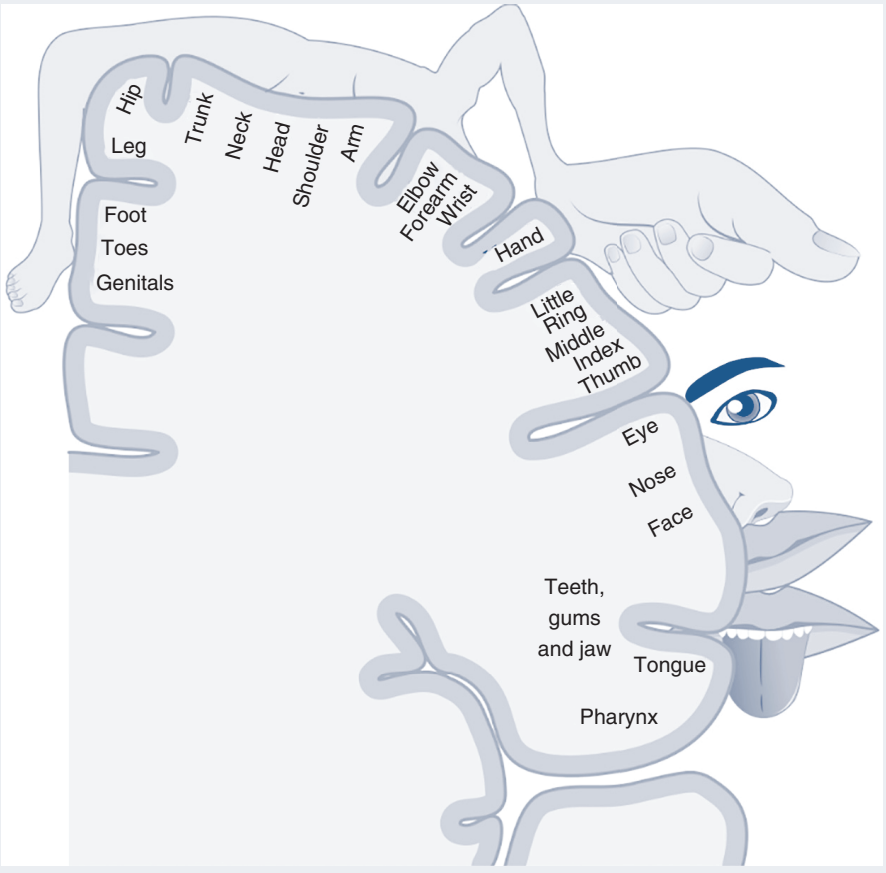
two halves, or **cerebral hemispheres**. Each hemisphere is roughly associated with a unique processing style (Sperry, 1985). The left hemisphere tends to be the seat of more analytic, serial, and logical reasoning. The right hemisphere tends to be the seat of more synthetic, parallel, and relational thought processes. Language function in most persons is localized to the left hemisphere, whereas spatial ability is usually concentrated in the right hemisphere. However, in about one fourth of left-handed people, this is reversed. Information is transferred back and forth between the hemispheres via a bundle of connecting axons that are known, collectively, as the **corpus callosum**.

A **fissure** is a large cleft or separation between two areas of brain tissue, whereas a sulcus refers to a smaller such separation. A **sulcus** typically separates neighboring ridges or folds of tissue. A single ridge of this kind is known as a **gyrus**. At a smaller scale, brain tissue is made up of nuclei (singular: nucleus), which are collections of cell bodies and tracts. The tracts are collections of nerve cell axons.

Figure 6.5 is a view of the cortical regions of the left hemisphere. The first thing you may notice is that the cortex has several large regions, called **lobes**. The **frontal lobe** is located to the anterior and is bounded by the central sulcus and the lateral fissure. It contributes to reasoning, planning, problem solving, language production, and related skills. The **temporal lobe** lies ventral of the lateral fissure. The temporal lobe contributes to auditory processing, language comprehension, pattern recognition, and visual object recognition. Just posterior to the central sulcus is the **parietal lobe**. It governs aspects of visual attention and visuospatial processing. The parietal lobe also plays a role in somatosensory processing (touch sensation). Finally, there is the **occipital lobe**, where visual information begins to undergo more extensive processing.

Two other cortical regions should be noted. Anterior to the central fissure is the precentral gyrus. It is the location of the **primary motor cortex**. Electrical stimulation of the primary motor cortex at a specific point provokes muscular contraction at the site of the corresponding body part. The function of the primary motor cortex is to plan and prepare actions, which subcortical regions and the spinal cord then convert into the precisely timed movements of particular muscle groups. Neurons in the primary motor cortex are laid out in a two-dimensional topographic map-like fashion, such that neurons near one another represent parts of the body that are near one another. There are large regions of neural tissue representing the hands and the face, and comparatively smaller regions representing the legs and the back. Due to this uneven distribution of surface area representation, it is referred to as a cortical "homunculus." If one were to sketch a humanoid creature with proportionately large hands and face along with proportionately small legs and back, it would look a fair bit like the weird homunculus creature that an evil wizard might grow in his laboratory. Just posterior to the primary motor cortex, on the postcentral sulcus, is the **primary somatosensory cortex**. Electrical stimulation of any portion of the primary somatosensory cortex triggers the perception of touch sensation coming from the corresponding part of the body. The primary somatosensory cortex processes sensory information arriving from the body surface. Much like the primary motor cortex, the primary somatosensory cortex is also arranged topographically like a homunculus (see Figure 6.5). Additionally, many brain regions in visual cortex also exhibit this **topography** in how they represent the locations of visual inputs (i.e., neurons near one another in visual cortex respond to photoreceptors in the retina that are near one another).

Figure 6.5 **A schematic midsection slice of the somatosensory cortex shows the large amount of cortical surface area devoted to the hands and face, with less surface area devoted to the trunk and legs.**



Source: OpenStax Collegederivative work: Popadius/CC BY.

The Split Brain

An important feature of the cortex is that information received from the right or left halves of the body is mapped onto the opposite, or **contralateral**, side of the brain. This means that touch input on the left side of the body sends signals through the spinal cord that end up in the right hemisphere of the brain in somatosensory cortex and vice versa for touch input on the right side of the body. Likewise, a visual stimulus presented in the

left visual field—everything to the left of where your eyes are pointed—is projected onto the right hemisphere, whereas a stimulus presented in the right visual field is projected onto the left hemisphere. Information presented to one side of the body is not processed on the same, or **ipsilateral**, side of the brain. This contralateral organization is the result of the crossing of fibers originating from sensory neurons that are located on one side of the body over to brain areas on the other side. The same holds true for motor control: The left hemisphere sends commands to the right side of the body, while the right hemisphere controls the left side.

A split-brain patient has had his or her corpus callosum surgically cut to prevent the spread of epileptic seizures. Experiments with these patients demonstrate just how specialized the two brain hemispheres are. In these studies, a patient sees either a word or a picture of an object that is presented to the left or right visual field. This presents the information about the object to either the right or the left hemisphere, respectively. The patient can identify the object either by naming it or by grasping the object the word represents from among a set of objects that are hidden from view. The word can be vocalized only if the information that has been presented to the patient reaches the left hemisphere, as the left hemisphere is specialized for language. Either hemisphere with information about the object can identify it through grasping because this response relies on tactile information alone (Gazzaniga, 2005).

Imagine that the word *spoon* is presented to a split-brain patient in his or her left visual field. The information projects to the right hemisphere. The patient, in all likelihood, cannot read the word because, in most persons, language ability is lateralized to the left hemisphere. He or she also cannot identify the object with his or her right hand because that is controlled by the left hemisphere, which has no knowledge of the spoon. The participant can, however, reach out and select a spoon with his or her left hand. Imagine another scenario. The word *spoon* is flashed in the right visual field and, so, projects to the left hemisphere. The participant can at this point say that it is a spoon, as well as grab a spoon with his or her right hand. Correspondingly, this patient would not be able to identify the spoon with the left hand.

THE NEUROSCIENCE OF VISUAL OBJECT RECOGNITION

You may recall the different models of pattern recognition we discussed in Chapter 4. These models were all ways of describing how we visually recognize objects. In this section, we examine several clinical disorders to see what they can tell us about the neurological basis of these recognition mechanisms. But first, we must introduce a basic principle of information processing in the visual system—the partitioning of visual input into separate streams.

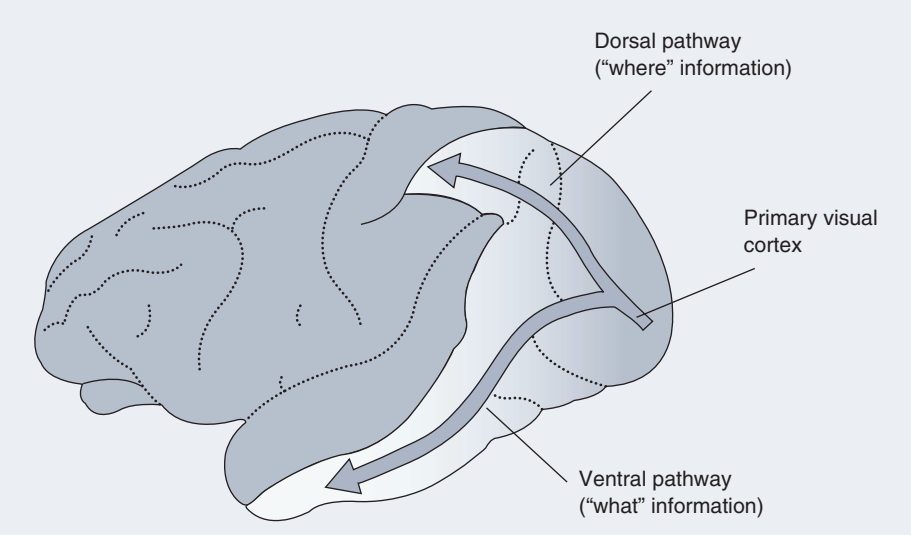
The visual system, as we mentioned earlier, breaks objects down into their parts or features, with different areas of the brain handling them. Research has revealed two distinct anatomical pathways for the processing of visual information (Ungerleider & Mishkin, 1982). Visual inputs undergo preliminary processing in the primary visual

cortex—located in the occipital lobes, at the back of the head. Following this, the information is divided and projects to two different parts of the brain. One stream, called the **dorsal visual pathway**, travels upward to the parietal lobe, where information about motion and location is extracted. Ungerleider and Mishkin originally called it the "where" pathway because of its representation of the spatial locations of objects. More recently, it has been called the "where/how" pathway because the parietal lobe (which includes the somatosensory cortex and is adjacent to the motor cortex) seems to code also for "how" the body would interact with the objects in those locations—such as catching a ball or pointing at an object (Goodale & Milner, 1992). A second stream, the **ventral visual pathway**, carries data about color and form and travels downward to the temporal lobe. It is referred to as the "what" pathway because of its role in object identification. Figure 6.6 shows the locations of both.

Visual Agnosias

Oliver Sacks is a neurologist who is also a gifted writer. In *The Man Who Mistook His Wife for a Hat,* he recounts the case histories of several patients with disorders so bizarre that they border on the fantastical (Sacks, 1998). Take, for example, the case of patient P., who is a music teacher. P.'s gaze is unusual; his eyes dart around as they take in Sacks's face, fixating on one feature and then another. P. is unable to recognize a rose and describes it as "a convoluted red form with a linear green attachment." During one interview

Figure 6.6 The dorsal and ventral pathways of the monkey's visual system.



with Sacks, P. took hold of his wife's head as he apparently tried to lift it and put it on his own head. Sacks writes, "He had . . . mistaken his wife for a hat!" Although it may seem improbable, such patients really do exist. They have a pattern recognition disorder called visual agnosia. What is wrong with these patients' brains? How can they make such errors? Let's take a look.

A **visual agnosia** is an inability to recognize a visual object (Farah, 1990). Visual agnosias are perceptual in nature and cannot be attributed to problems that have to do with memory. They are modality specific. In one case study, a woman who was not able to recognize a picture of a hammer was able to recognize an actual hammer through the use of another sense. She was able to pull a hammer out of a bag that was filled with objects. Furthermore, it seems that in patients with visual agnosias, most other visual system–processing abilities remain intact. These patients do not have difficulty detecting motion, interpreting location information, or reaching for handles of objects such as a coffee mug. Agnosias are associated with damage to brain regions in the ventral visual pathway that carry out visual object recognition.

There are two broad categories of visual agnosia. Persons with **apperceptive agnosia** cannot assemble the parts or features of an object into a meaningful whole. Persons with **associative agnosia** perceive this whole but have difficulty assigning a name or label to it. The apperceptive form of the disorder seems to involve a disruption of the formation of an object representation and is, therefore, more perceptual, or "lower level," in nature. The associative form involves a disruption of the ability to categorize or identify objects and can be considered more cognitive, or "higher level," in nature.

Apperceptive Agnosia

In patients suffering from apperceptive agnosia, most of their basic visual functioning is intact. This includes the capacity to see details, discriminate between lights of different brightness, and perceive color. However, they have great difficulty naming, matching, copying, or telling the difference between simple visual forms. In one test, patients were asked to indicate which of four drawings of common objects matched a single target drawing. The apperceptive agnosic patient, Mr. S., was unable to do this. He matched a circle to a triangle and a paperclip to a key. Mr. S. was also incapable of copying letters of the alphabet. While attempting to copy a capital letter *X*, he drew two separated oblique lines.

So how can we explain this disorder? The current explanation is that apperceptive agnosics have suffered a disruption of their perceptual grouping mechanism. In the psychological approach chapter, we described grouping as the process of assembling basic perceptual features. These agnosics lack the ability to integrate the various features of an object, although they can perceive the individual features accurately. For example, an apperceptive agnosic patient might be able to recognize that a picture of a car has wheels, windows, and doors, but he or she may not be able to combine these parts into a complete percept of a car (see Figure 6.7). These individuals have sustained general damage to the occipital lobes and nearby areas. In these individuals, this kind of damage interferes with the combination or assembly of features into a unified object representation.

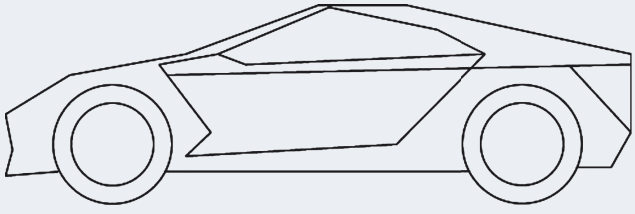
Another disorder that may be considered a type of apperceptive agnosia involves difficulty recognizing objects when they are viewed from unusual angles or are lit unevenly.

This disorder sometimes goes by the name of **perceptual categorization deficit**. It often goes unnoticed, as patients who have the disorder usually have no problems performing other visual tasks. The existence of the disorder can be revealed under certain testing conditions, however. Affected persons are, for instance, not able to say that the object represented in a picture of an upright ladder is the same object as that represented in a picture of a ladder as viewed from below. Notice that this deficit constitutes a failure of object constancy as it is an inability to recognize an object after it has undergone a transformation, such as a change in perspective (see Figure 6.8). These patients have sustained damage to the right hemisphere, especially the right parietal lobe.

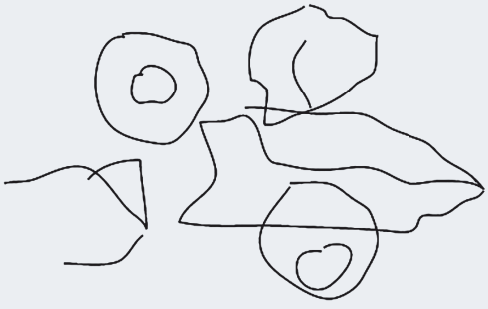
Associative Agnosia

Associative agnosias have three diagnostic criteria. First, affected persons have difficulties recognizing objects visually. Second, these patients can recognize objects using sensory modalities other than vision—for example, touch or sound. Third, they do have

Figure 6.7 A patient with apperceptive agnosia can recognize the features of an object but, when asked to draw it, is unable to put the features together in the right way.

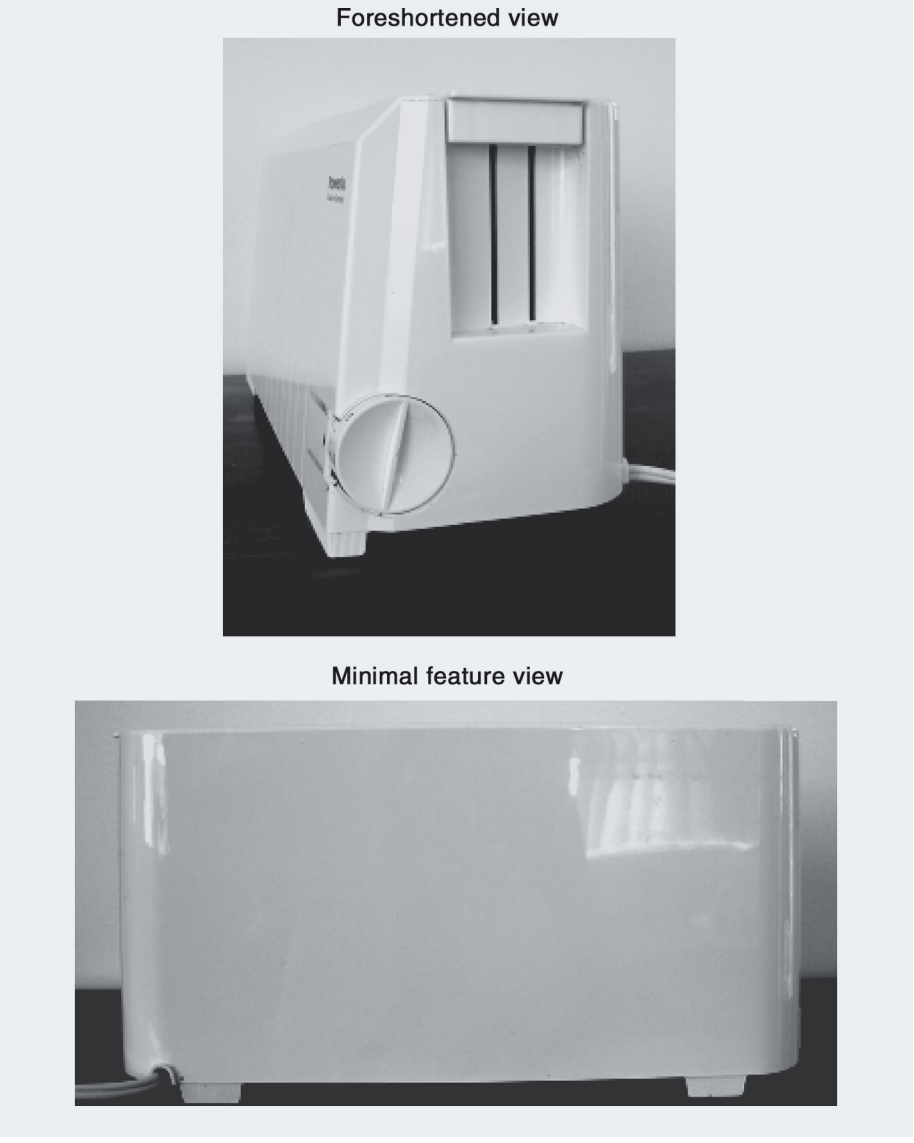


Picture of a car



Drawing of same car by apperceptive agnosia patient

Figure 6.8 Patients with visual agnosia can have difficulty recognizing an object when it is viewed from an unusual angle (foreshortened view) or from a perspective with few features visible (minimal feature view).



the ability to perceive objects holistically, at least in the operational sense of their being able to copy or match drawings of objects. It is with respect to this third criterion that associative agnosics differ from apperceptive agnosics, who cannot perform these tasks.

Individuals with this disorder demonstrate a behavioral anomaly. They can copy a drawing accurately, albeit quite slowly, but cannot name what it is they just copied. One patient, L. H., copied line drawings of a teabag, a diamond ring, and a pen rather precisely but could not supply the correct verbal label for any of the items (Levine & Calvanio, 1989). This implies that associative agnosics are able to perceive entire objects correctly. Their difficulty lies in coming up with a name for what they see.

Several theoretical explanations of this disorder have been proposed (Farah, 1990). Early explanations posited a disconnection between language areas that contain linguistic representations and visual areas. Under these theories, the two areas have remained intact; it is the connections between them that are damaged. This prevents them from matching visual inputs to verbal labels. Humphreys and Riddoch (1987) later elaborated on this idea. They believed that the deficit was due to damage to a system that contains stored visual object representations. More recent explanations use neural network pattern recognition architectures (Behrmann & Plaut, 2013). Damage to the connections that enable the formation of object representations can then account for the observed anomalous behaviors.

Face Perception

Prosopagnosia. **Prosopagnosia** is an inability to recognize and discriminate faces, despite the capacity to recognize other types of visual stimuli and the presence of generally intact intellectual functioning. Prosopagnosia is considered a type of associative agnosia because in persons who have the disorder, the ability to *perceive* faces is mostly intact. The difficulty lies in being able to *identify* the faces. Individuals with this disorder are sometimes unable to recognize close friends, family members, and, in some cases, even the reflection of their own face in the mirror (Burton, Young, Bruce, Johnston, & Ellis, 1991; Parkin, 1996). The ability to discriminate faces is also impaired. Prosopagnosics have problems telling faces apart from one another.

Research with monkeys has uncovered the existence of cells that respond selectively to faces. The area that has been identified is in the inferotemporal (IT) cortex, part of the ventral pathway that is responsible for the recognition of forms (Desimone, Albright, Gross, & Bruce, 1984; Perrett, Rolls, & Caan, 1982). Bruce, Desimone, and Gross (1981) found that some IT neurons fired most rapidly when the animal was presented with the stimulus of a complete face, either that of a monkey or that of a human being. The neurons fired less rapidly when the monkeys were presented with images of incomplete faces—those that had one feature, such as the eyes, removed. A caricature of a face produced even slower responding. Wachsmuth, Oram, and Perrett (1994) also measured neural firing in area IT in monkeys. They discovered that there were cells that became active when the animal was shown a picture of a face or a face and a body. Response

dropped significantly when the stimulus consisted of a picture of a body only, with the face covered. More recently, Rodrigo Quian Quiroga (2012) was able to record from individual neurons in the hippocampus of humans (during brain surgery) and found evidence for **sparse distributed coding**, whereby a very small set of neurons are highly selective in their response to a particular person's face, such as a celebrity (or even to the written name of that celebrity). Quiroga refers to these cells as "concept cells" and stresses that each one participates in a small assembly of cells in the hippocampus that cooperate to form a conceptual representation.

Neural Coding. It is worth digressing for a moment to discuss the idea of neural coding. The research just described makes it seem possible that we might have individual neurons that respond to individual faces—that single cells fire only when the stimulus consists of a particular face. This type of representation is known as **localist coding**. In localist coding, a single cell fires in response to the presence of a particular face—say, your grandmother's—but not in response to any other face. Another cell would fire in response to your grandfather's face but not to your grandmother's or anybody else's. Although localist coding of this sort is possible, it is unlikely for two reasons (Goldstein, 2002). First, it is now known that many cells that respond to one face often respond to others as well. Second, a tremendous number of cells would be required for the coding of all the faces we know and of all the possible views and expressions of these faces.

A more likely explanation is some form of **distributed coding**, in which a specific face is coded for by a specific pattern of activation among a group of cells. In a *widely* distributed coding scheme, there would be a large number of cells that cooperate to encode the representation of a particular face or a particular concept. However, in a *sparse* distributed coding scheme, your grandmother's face might cause cells A and B to fire rapidly, cells C and D to fire slowly, and cells E and F to fire at a moderate rate. Your grandfather's face then might induce a similar pattern of activation among a few of those cells but would also have some other cells that are specific to it. Sparse distributed coding can efficiently represent a large number of faces or facial attributes and in a way that requires fewer neurons.

In human beings, "face cells" seem to be located in the **fusiform face area (FFA)** and in the medial temporal lobe (which includes the hippocampus and related areas). Studies using fMRI show that pictures of faces activate FFA, located in the fusiform gyrus of the human IT (Clark et al., 1996; Kanwisher, McDermott & Chun, 1997; Puce, Allison, Asgari, Gore, & McCarthy, 1995). This region appears to be dedicated to the processing of face information. That there should be a part of the brain devoted to faces is not surprising, considering the important role they play in social interaction. Evolutionary forces may have selected for a face-processing mechanism that differs functionally from how other objects are recognized, although there is an ongoing debate over whether faces constitute a special class of perceptual stimulus. For instance, FFA is also specifically active when automobile experts classify images of cars (Tarr & Gauthier, 2000).

INTERDISCIPLINARY CROSSROADS: PERCEPTUAL BINDING AND NEURAL SYNCHRONY

If you think back to Chapter 4 and the feature integration theory of pattern recognition, you will remember that serial search takes place when the target to be identified shares several properties with the distractors. The only way to locate it successfully is to focus attention on it to determine whether the particular shape and color at a given location match. Attention in this case was said to "glue" the different features together, cementing in place if you will the shape of a letter *T* and the color blue at a specific location in the visual field.

How does this happen? The parts of the brain that code for shape, color, and location are actually in different areas. Shape is processed by neurons in the IT cortex of the visual pathway in the temporal lobe, color is coded for by neurons in a location called V4 in the occipital lobe, and location is coded for in the parietal lobe, part of the dorsal processing stream. How does all this information come together to produce a unified perceptual object? This question of how to recombine the various features of an object is called the binding problem.

One group of researchers appears to have come up with a solution (Engel, König, Kreiter, Schillen, & Singer, 1992; Singer, 1996). They suggest that an object is represented by the coordinated and phase-locked activity of a constellation of cells—a concept known as **neural synchrony**. Distinct cell subgroups of this constellation stand for individual features and may be separated by relatively large physical distances in the brain, but the synchronized firing rates of all of them serve to represent an entire object. The same group of neurons that code for a particular feature in one object can participate in different cell assemblies and thus stand for this same feature in other objects. For example, a network of cells representing the color red can participate in the synchronized assembly for a stop sign. These same cells can then participate in the synchronized assembly for a tomato.

A problem with the existence of functionally distinct cell populations in the brain is how they stand out amid all other ongoing activities. Singer (1999) proposes three solutions. First, other neurons not participating in the assembly can be inhibited. Second, the amplitude or strength of the cells in the assembly can be increased. Third, the cells in the assembly can synchronize their firing rates. This temporal synchrony means that all the participating neurons would fire at the same time. It is helpful to use an analogy here. Imagine a group of drummers in a band. If all the drummers banged out different rhythms, they wouldn't stand out much. But if the other musicians played more softly and the drummers played louder and, most important, began beating the same rhythm, we would hear the drums as a salient ensemble.

There are some problems with temporal synchrony (Singer et al., 1997; von der Malsburg, 1981). Not all cells coding for features fire at the same rate. To overcome these difficulties, it was postulated that single or partial discharge rates of neurons, rather than their entire pattern of activity, can be synchronized (Gray, König, Engel, & Singer, 1989; Singer & Gray, 1995). In this scheme, only one cell output or a small

(Continued)

(Continued)

number of outputs that are part of a given sequence of outputs are timed to coincide with others. Returning to our musical example, it would be like having every fourth beat of one of the drummers coinciding with every fourth beat of the other drummers. Alternatively, two drummers could synchronize recurrent clusters of beats. This partial discharge synchrony would allow the drummers to play at different rates yet still coordinate their activity.

There is evidence to support the synchrony hypothesis. Engel et al. (1992) measured the electrical activity of spatially separated cells in the primary visual cortex. These cells respond individually to lines of a given orientation (Hubel, 1982; Hubel, Wiesel, & Stryker, 1978). In this case, both cells responded to vertical lines. When each cell was stimulated individually with a vertical line, the cells showed moderately correlated activity. But when a single prominent vertical line was used to stimulate both, they showed much stronger coordinated activity. They fired short-lived bursts of output at the same time, alternating with periods of lowered activity. These two cells firing out of synchrony suggest that they represent object parts. When they fire synchronously, they could then be part of a larger assembly that codes for the entire object. Neural synchrony could be the basis for more than just perceptual binding. It may explain memory formation and consciousness as well (Axmacher, Mormann, Fernandez, Elger, & Fell, 2006).

More recent research on this topic shows that the brain can coordinate far-flung activity in a much more flexible way than previously assumed. Shadlen and Movshon (1999) have suggested that precise neural synchrony may not be necessary for linking together different distributed codes. Guttman, Gilroy, and Blake (2007) differentiate between temporal synchrony and temporal structure. In the first case, neurons must be precisely in phase with one another. The impulses or spikes representing action potentials have to occur at the same moment, either for the entire pattern or for subsets within the pattern. In the second case, there can be similar patterns of impulses over time in either location, and these can be linked. Imagine that neurons in one brain area fire two bursts of high-frequency impulses followed by one burst of low-frequency impulses. In another area, this same pattern occurs but a second later. Even though the impulses don't happen at exactly the same time, their similar patterning or temporal structure can still serve as a basis for linking together the information coded by the two different cell assemblies.

THE NEUROSCIENCE OF ATTENTION

Attention plays a key role in many different cognitive phenomena. This may explain why there are at least six distinct brain structures that underlie attentional effects (Posner & Peterson, 1990). We can think of these structures as being interconnected and as collectively forming an attentional neural network. Figure 6.9 shows their locations. These different areas work, sometimes in concert, to control attention.

We will describe the location and function of each of these structures in detail in the numbered list that follows, but first, let us give you a sense of how they work together.

The **reticular activating system (RAS)** is responsible for our overall arousal level, the extent to which we will pay attention to anything. The thalamus regulates the amount of incoming sensory information that reaches the cortex for processing. A more refined processing of the information becomes possible once it reaches the cortex. This processing becomes the job of the parietal lobe, which allocates attentional resources to certain aspects of the information, such as an object's spatial location. Ultimately, the cingulate cortex initiates a response on the basis of what is attended.

The following structures are considered parts of the neural attention network:

1. The RAS, with cell bodies located in the hindbrain, consists of a network of several dozen neural circuits. These neural circuits have projections to diverse areas of the cortex. The function of this system is very basic: It controls the brain's overall arousal and alertness levels. Activation of the RAS is linked to the ability to sustain attention over time. Bilateral lesions to the RAS result in a comatose state in which patients are unresponsive to most stimuli.

2. The **superior colliculus** is located in the midbrain. Its function seems to be the shifting of visual attention from one object or location in the visual field to another. Monkeys with superior colliculus lesions are unable to shift their gaze to a novel stimulus that appears suddenly in the peripheral visual field (Schiller, Sandell, & Maunsell, 1987). This type of deficit has also been observed in humans in whom there has been a degeneration of the colliculus (Rafal, Posner, Friedman, Inhoff, & Bernstein, 1988).

3. The **thalamus** is dorsal and anterior to the superior colliculus. It receives inputs from the RAS and forwards them to the cortex. It is a part of the general RAS arousal circuit. The thalamus also serves as a relay center. It forwards incoming sensory information to the different parts of the cortex specialized for processing them. It is believed that a particular neural circuit within the thalamus is responsible for regulating those sensory messages. It serves as a filter or gatekeeper, determining how much information is allowed to continue to the cortex for further processing. A study that used PET observed a greater amount of thalamic activity in instances when a participant was asked to single out a target item from a collection of items, as compared with when the participant was asked to identify a target item that had been presented by itself (LaBerge & Buchsbaum, 1990).

4. The parietal lobe seems to service a broad variety of visual attention-related tasks. In monkeys, cells in the intraparietal sulcus of the parietal lobe are active when a particular spatial location is attended (Colby, Duhamel, & Goldberg, 1996). In humans, the **intraparietal sulcus** is active in processing a combination of visual stimulus features, such as color and shape, which suggests that it is responsible for the binding together of features in visual search (Wojciulik & Kanwisher, 1999). The parietal lobe is also the site where attentional resources are allocated to different tasks (Coull & Nobre, 1998; Fink, Dolan, Halligan, Marshall, & Firth, 1997). Think back to Kahneman's

capacity model of attention from Chapter 4. You may remember that there was a component of that model called the allocation policy. Its job is to distribute attentional resources among a range of possible activities. The parietal lobe may be the neural equivalent of that component.

5. The **cingulate cortex** is believed to be part of the network of brain regions in which a response is selected, especially in cases where the response entails the inhibition of an alternative response. An example of the selection of a response in the presence of competing inputs is the Stroop effect (MacLeod, 1991). In this phenomenon, participants are asked to look at a color word (e.g., *red*) printed in a particular color (e.g., red ink or blue ink). They then have to name the color of the ink. These two attributes can be congruent, as when the word *red* is presented in the color red. They can also be incongruent, as when the word *red* is presented in the color blue. In the congruent condition, observers react quickly because the two attributes prompt similar responses. In the incongruent condition, observers' reactions are slowed because one of the attributes elicits a response that must be inhibited, which takes additional time. Milham et al. (2001) found activation of the cingulate cortex when participants were responding to the incongruent condition of the Stroop task.

6. We already have introduced the frontal lobes. With respect to attention-related tasks, their job—like that of the parietal lobes—is quite varied. The frontal lobes play a role in the selection of motor responses, in goal-directed action, and in adherence to a task, among other things. The frontal lobes feature prominently in executive function and problem-solving situations, so we reserve the greater part of our discussion of the frontal lobes for later.

Neural Models of Attention

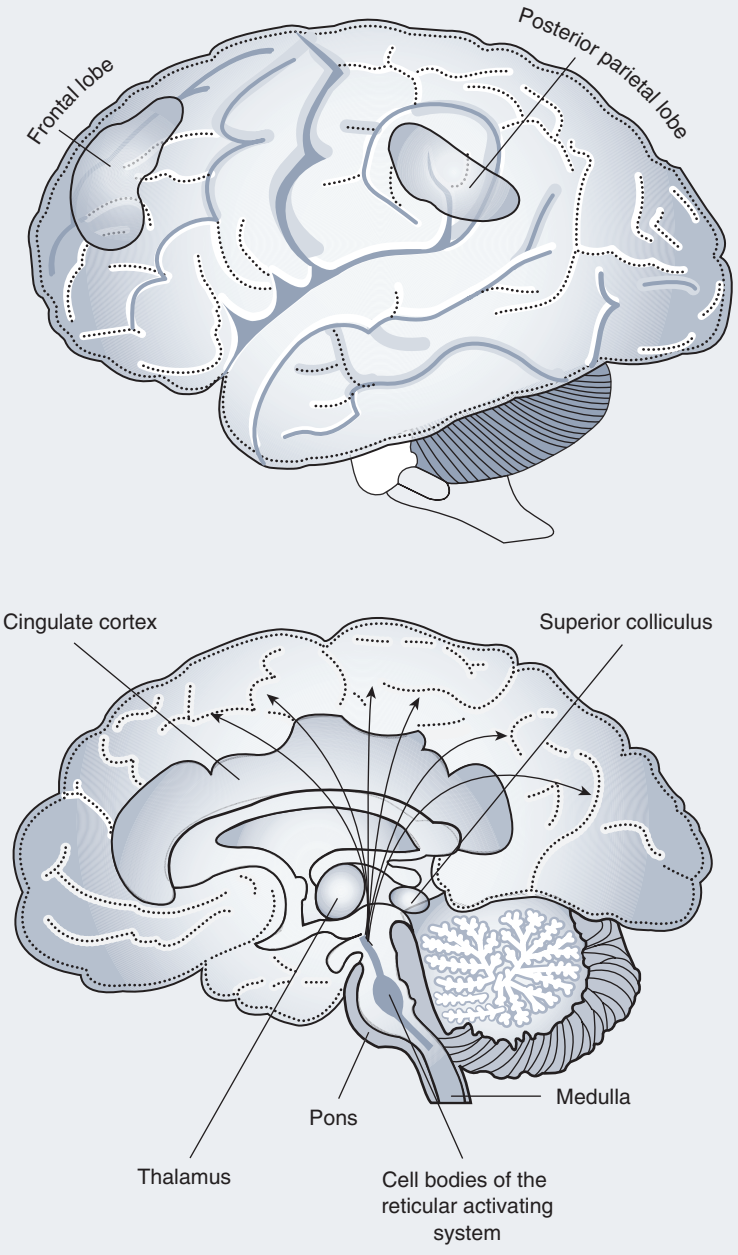
Banich (2004) outlines several types of neural models of attention. We describe two of them here. In component process models, distinct brain areas each have a unique and nonredundant function. In distributed network models, the areas can be characterized as having some operational overlap.

A Component Process Model

Posner, Inhoff, Friedrich, and Cohen (1987) propose that each of the multiple brain areas responsible for the control of attention performs a distinct operation. Their model specifically describes the changes that occur in visual selective attention, where attention is shifted from one spatial location to another. They posit that cells of the parietal lobe are used to disengage attention or remove it from a specific location. The superior colliculus then moves attention to a new location. Finally, the thalamus engages the attention and focuses it on the new location.

The data that support this model come from an experimental paradigm in which participants are asked to focus on a central cross and are then cued to respond to a target, such as an asterisk, that can appear on either side (Posner, 1980). The target appears in

Figure 6.9 Brain structures involved in the function and control of attention.



one of two boxes to the left or to the right of the cross. There are two types of trials. On "valid" trials, the box that will contain the target briefly lights up. On "invalid" trials, the opposite box without the target lights up. For one of the conditions, the majority (80%) of the trials are valid while the remaining trials (20%) are not. Participants in this condition respond faster to valid trials because the cue is a reliable indicator of where the target will appear. In effect, they are able to move a visual "spotlight" of attention into the target location before it appears and respond to it more rapidly. On invalid trials, it takes longer, ostensibly because they move the attentional spotlight in the opposite direction and then must bring it back to the target location.

In a variation of the Posner visual selective attention paradigm, patients with damage to either the right or left parietal lobe had no difficulty responding when the cue and target appeared on the side contralateral to the damage. In these cases, the cue was valid, and the damaged parietal lobe that processed information coming from the opposite visual field did not have to disengage attention. But during invalid trials in which the cue appeared on the side ipsilateral to the lesion and the target appeared on the other side, response time was slow. In these cases, the parietal lobe of the damaged hemisphere needed to disengage attention and had difficulty doing so (Posner, Walker, Friedrich, & Rafal, 1984).

Patients with damage to the superior colliculus show delayed responses in valid trials. This suggests that they are slower to shift attention from the fixation cross to the cued target location, no matter which side of the cross it is on (Rafal et al., 1988). Individuals with sustained damage to the thalamus exhibit yet another result. They are slower to respond whenever the target appears on the side contralateral to the side of the thalamus that is damaged, whether the trial is valid or invalid. This implies that they cannot engage attention on the target.

Distributed Network Models

Mesulam (1981) proposes an alternative model for the neural machinery that controls attention. In this model, the separate neural structures are not quite as specialized and functionally independent as they are in the component process model. Instead, the functions of the different areas overlap to some degree. Each brain region performs a major operation that is attention related but can perform other attention-related functions as well. This redundancy in the network implies that any given area can suffer damage, while the system as a whole will maintain some of the functionality subsumed by the damaged region.

This model explains how attention is directed to areas outside the body and is derived from the deficits observed in specific types of brain damage in monkeys and humans. The model encompasses four brain areas, each playing a primary, but not exclusive, role in controlling an aspect of attention. First, the posterior parietal cortex provides a sensory map or representation of the space in which attention will be directed. Second, the cingulate gyrus in the limbic cortex plays a motivational role. It determines what should be paid attention to and what can be ignored. Third, the frontal cortex coordinates motor programs for attention-related actions. These actions include fixating on certain regions in the visual field, scanning across the visual field, or reaching out to grasp an object. Finally, reticular structures generate arousal and vigilance levels.

Let's give an example that will illustrate how all these areas might work together. Imagine that Susan is sitting at the breakfast table. She missed dinner the night before and so is quite hungry. Her posterior parietal cortex contains a topographic sensory representation of the breakfast table that includes the locations of items such as a glass of orange juice and a bowl of cereal. Objects that are relevant for her current goals may have neural representations in parietal cortex that are more active than those for irrelevant objects (Behrmann, Geng, & Shomstein, 2004). In many cases, the shifting of attention may simply be an emergent result of these neural representations competing against one another until certain ones become strongly active while others become less active—as with Desimone and Duncan's (1995) biased competition model from Chapter 4. Her cingulate cortex directs her attention to items and/or regions of significance. In this case, she would pay more attention to the orange juice than to the tablecloth. Her frontal cortex supervises motor actions, such as her reaching out to pick up the orange juice. Underlying all this are arousal levels that have been determined by her reticular system. If Susan had a good night's rest, her arousal and vigilance ought to be at high levels.

If multiple brain areas are responsible for our attention, then you might be wondering how these areas coordinate their activity. This might be achieved via a synchronization of nerve cell firing in each area. To learn more about how such synchronization might take place in perception, see the "Interdisciplinary Crossroads" section.

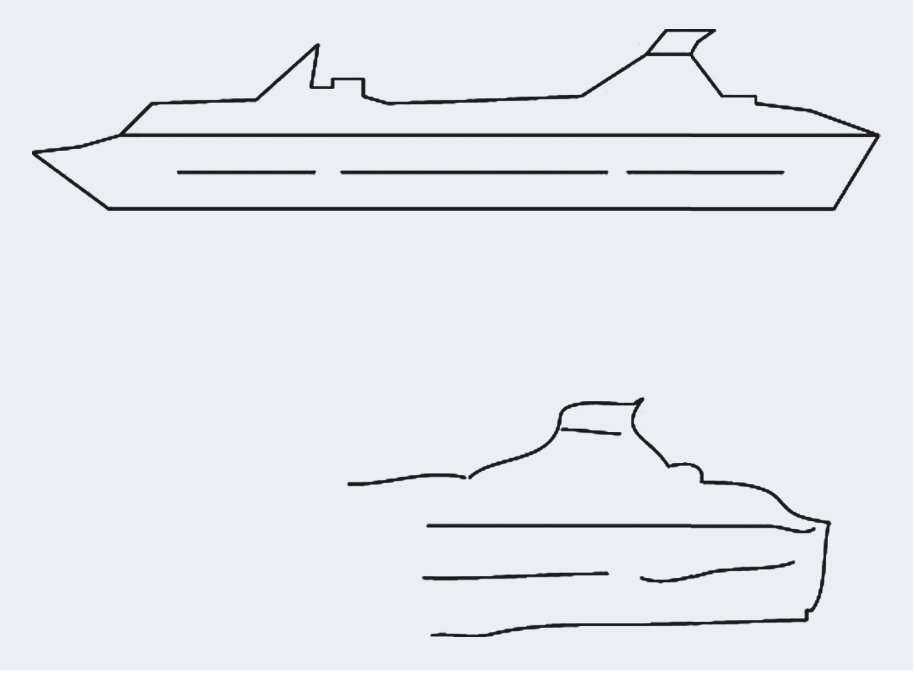
Disorders of Attention

Hemispatial Neglect

Neglect is the opposite of attention. It refers to information that we ignore or fail to pay attention to. Although we do notice a lot of what we perceive, just think of all that we miss. In the case of vision, we can usually see what we are looking at directly but not what is in the periphery. This is normal and doesn't usually impede everyday functioning. But certain types of brain damage can result in a profound loss of attention. In **hemispatial neglect**, there is a tendency to ignore one side of the body and its surroundings. These include visual, auditory, and touch information. This disorder occurs with damage to the superior temporal gyrus usually in the right hemisphere (Karnath, Ferber, & Himmelbach, 2001). This gyrus receives inputs from the auditory system and from the dorsal and ventral visual systems and has outputs to the basal ganglia, so symptoms of damage to this area include neglect for each of these senses as well as motor impairments.

Remarkably, patients with left hemispatial neglect can perceive only the right side of objects (Behrmann & Geng, 2002; Driver & Mattingley, 1998). When asked to recall what they've seen, they describe only the right side, and when asked to draw what they see, they copy only the right side (see Figure 6.10). However, they also have difficulty attending to two items that are presented at the same location. This is true visually when two objects are at the same location, but it is also true for audition. When two notes at different pitches are played at the same time, patients have difficulty telling which tone was played first (Cusack, Carlyon, & Robertson, 2000). With training, some of these deficits can be overcome. For example, patients can be cued in various ways to attend more to one side (Frassinetti, Pavani, & Làdavas, 2002). It is important to point out that

Figure 6.10 A patient with left hemispatial neglect, when asked to draw a picture of a ship, can reproduce only the right side.



Source: Adapted from Shallice (1982).

neglect is an attention problem, not a perceptual one. These patients have intact sensory systems. They can see and hear perfectly well but are unable to allocate attention to the perceptual information.

Attention-Deficit Hyperactivity Disorder

Commonly known as ADHD, **attention-deficit hyperactivity disorder** is characterized by distractibility, impulsivity, mood swings, short temper, overreaction to small stressors, and difficulties in planning or carrying out plans (Wender, Wolf, & Wasserstein, 2001). This disorder has been estimated to afflict around 10% of children in the United States as well as a smaller percentage of adults. A complicating factor is that the disorder is difficult to diagnose. Some children may be misdiagnosed with ADHD. Likewise, some children with the disorder may be misdiagnosed as not having it and are considered to be on the high end of the normal spectrum for these symptoms.

There is clear evidence for a genetic basis to ADHD. It runs in families. The gene for one type of dopamine receptor is more common in some patients (Faraone, Doyle, Mick, &

Biederman, 2001). They also have slightly smaller than normal brain volumes with smaller cerebellums and a smaller right prefrontal cortex (Giedd, Blumenthal, Molloy, & Castellanos, 2001; Ravizza & Ivry, 2001). However, these differences are small, and brain imaging cannot be used to make a diagnosis (Kalat, 2004). The current drug of choice for treating ADHD is a stimulant medication called Ritalin. In children, Ritalin increases the ability to concentrate, reduces impulsivity, and improves both school grades and social interaction (Elia, Ambrosini, & Rapoport, 1999). Similar results have been obtained with adults. Ritalin acts by boosting the action of the neurotransmitter dopamine.

THE NEUROSCIENCE OF MEMORY

Early research on memory asked a very basic question: Where in the brain are memories located? Karl Lashley (1950) was one of the first researchers who attempted to find out. He was searching for the **engram**, a physical change in a part of the brain that is associated with learning. His methodology was to train rats to perform different tasks. For example, he would teach his rats how to find the solution in a maze, with the expectation that the animals would form a memory of the task. He would then introduce lesions to parts of the brain to see if the memory of the task was destroyed in the process. Lashley's rationale was that if you destroy the part of the brain that contains the engram for the task, the rat's memory of how to perform it will be lost. In one experiment, he destroyed larger and larger areas of cortex in a graded manner. He found that the greater the amount of tissue that had been destroyed, the greater was the amount of time needed to retrain the rats to perform the task. But in no case did the animals actually forget how to perform the task.

Lashley concluded that memories are not laid down in any one area of the brain but that all parts of the brain participate in memory storage—a principle referred to as **equipotentiality**. This principle has some interesting implications. It suggests that multiple copies of a single memory are stored throughout the brain and that multiple brain regions participate in memory formation. We now know that there is some degree of truth to this, but different brain regions do appear to have somewhat different functions for memory. The distinct cortical areas that process information derived from a single sensory modality do contain very different representations. It is also the case that many brain areas are involved in memory. These areas include the hippocampus, the entorhinal cortex, the amygdala, the striatum, the left parietal cortex, and the prefrontal regions (see Figure 6.11). In this section, we will examine some of the more recent work on the neuroscience of memory. Our focus, as before, is to describe the different brain structures that underlie different types of memory processes and to formulate models of how they work together to create memories.

Learning and Memory

Learning takes place when an event provokes a nervous system alteration that in turn provokes a change in an organism's behavior. It is this change in the nervous system that is a memory. From a neuroscience point of view, a memory is essentially the formation

of a neural circuit—the forging of a pathway among neurons. The memory can then be considered to be the circuit itself. When the circuit is reactivated at some point in time after its establishment, the memory that corresponds to the circuit is experienced and informs the organism's behavior. In this way, nervous systems preserve the past and help organisms adapt to novel environments.

Learning at the cellular level requires **synaptic plasticity**, a capacity for change in the structure or biochemistry of a synapse. Donald Hebb (1949) was perhaps the first to understand this and formulated what has become known as the **Hebb rule**. It states that if two connected neurons are active simultaneously, the synapse between them will become strengthened. An easy catchphrase for this phenomenon is "Neurons that fire together wire together." Imagine two cells, A and B. An increase in the strength of their connection means that when Cell A is active, it is now more likely to activate Cell B. Cells A and B have now become linked together in a circuit: When A fires, B fires also.

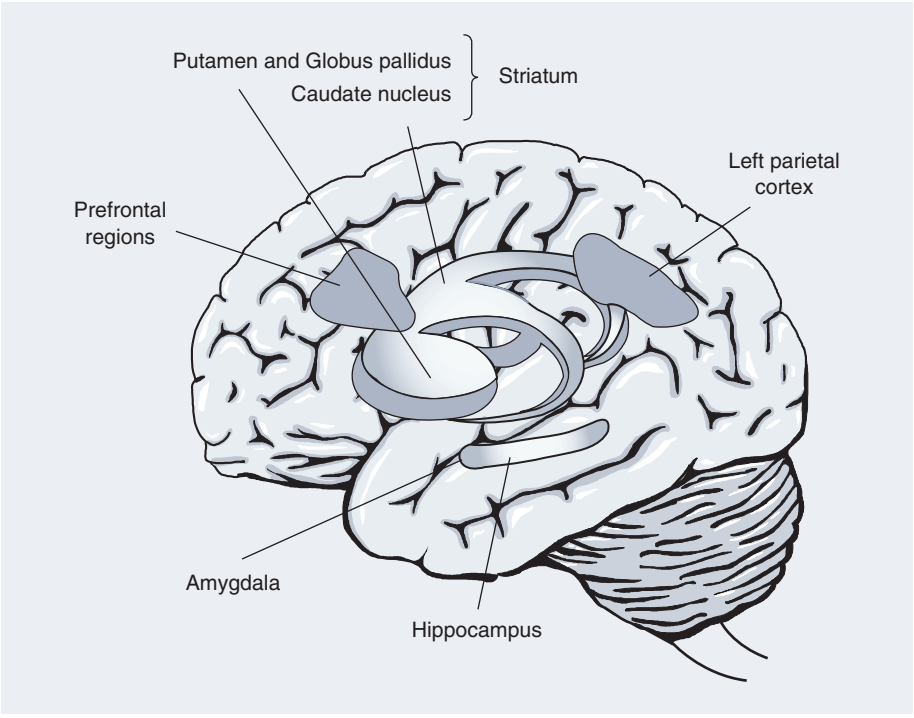
The specific neural change that mediates this Hebbian form of learning is called **long-term potentiation (LTP)**. It is the enduring facilitation of synaptic transmission that occurs following activation of a synapse by intense high-frequency stimulation. Rapid repeated stimulation is the key to LTP. It induces changes such as an increase in the number of receptors (Tocco, Maren, Shors, Baudry, & Thompson, 1992), as well as other structural alterations that result in increased neurotransmitter release (Edwards, 1995). These changes strengthen the synapse, increasing its responsiveness to further stimulation. LTP occurs in parts of the brain where learning and the creation of new neural circuits are important. It has been most often studied in the cells of the hippocampal system, which we will now discuss.

The Hippocampal System

We begin our discussion of the hippocampus with a famous case study (Corkin, 1984). H. M. was a patient who suffered from **epilepsy**. This is a disorder in which neurons fire uncontrollably, producing muscle spasms and seizures. H. M. did not respond to anticonvulsant medication. To stop his seizures, surgeons took drastic measures. They removed portions of his medial temporal lobes, including the hippocampus. The surgery was successful in the sense that it put an end to his seizures, but it left him with a profound memory deficit. After the operation, H. M. lacked the ability to learn any new information. He could have a conversation with a friend, turn around to talk to someone else, and in just a minute or so completely forget that his friend was even there. The removal of the medial temporal lobe tissue disrupted the process by which information is transferred from working memory to long-term memory. This process is called **consolidation**.

Despite his memory deficit, many of H. M.'s other memory abilities were unaffected. His working memory and problem-solving capacity were intact. In fact, he enjoyed solving crossword puzzles—but he would sometimes solve the same puzzle several times, as he forgot that he had already completed it! The content of his long-term memory prior to the surgery was also unaffected. This was true for both declarative and procedural knowledge. He could remember the location of the house he lived in and could still perform the skills he had acquired prior to the surgery. The case of H. M. demonstrates

Figure 6.11 Brain areas that form part of the human memory system.



that different memory functions are handled by different neural structures. Damage to one of those structures can selectively remove a component of memory ability. This is a theme that is repeated later in this section.

Hippocampal damage is associated with **anterograde amnesia**, an inability to retain new information following the traumatic incident that caused the damage. H. M. suffered from this type of amnesia, as he could not retain any new information after the surgery. As we will discuss shortly, it is the hippocampus and associated areas that are responsible for the consolidation of newly acquired information. Anterograde amnesia needs to be distinguished from **retrograde amnesia**, an inability to remember information acquired prior to the damage-inducing event. Retrograde amnesia is often caused by head injury, and in most cases, the loss extends only as far back as a week or more before the injury.

The job of the hippocampus is essentially to take information about "current events" that it has received from the cortex, process this information, and send it back out to the cortex, where it is stored in a more permanent state (McClelland, McNaughton, & O'Reilly, 1995). Inputs from the cortical regions contain representations of episodes or events that the organism has just experienced. More specifically, we may think of the hippocampus as an integrator. If the job of the cortex is to "divide and conquer" sensory

input into distinct processing streams, the job of the hippocampus is to "combine and reunite" them. The hippocampus weaves the disparate parts of an experience into a single unified memory. Let's use the example of a birthday party to illustrate this. The various parts of the cortex have divided your experience of the party into fragments. The visual system maintains an image of the presents; the auditory system, the sounds of your friends' singing; the gustatory system, representations of the taste of the cake; and so on. All these representations converge as inputs to the hippocampus, which combines them to form a whole memory of the birthday party in a sparse distributed code that retains its connections to the sensory cortices as well.

While Quiroga (2012) has single-cell recording evidence for sparse-coding representations of concepts inside the hippocampus, Alex Martin (2007) has fMRI evidence for those hippocampal representations maintaining their connections to the feature-based representations in sensory and motor cortices. Thus, when you remember a recent birthday party, the process may start with activation of a sparse distributed code in hippocampus, but it quickly spreads to activating the associated neural representations in visual, auditory, and gustatory cortices (Kiefer & Pulvermüller, 2012; Simmons, Martin, & Barsalou, 2005). In fact, even an abstract concept, like *justice*, appears to have a relatively stable pattern of activation in left hemisphere brain areas, including some activation of sensory and motor cortices (Binder, Westbury, McKiernan, Possing, & Medler, 2005; Borghi et al., 2018; Vargas & Just, 2020).

Neural Substrates of Working Memory

In Chapter 5, we discussed the psychological components of working memory. These are the articulatory loop, where operations on verbal representations are performed; the visuospatial sketchpad, where visual information is processed; and the executive control system, which coordinates activities. Recent neuropsychological evidence suggests that these psychological constructs are indeed functionally distinct units and may even be identified with particular brain regions.

Smith, Jonides, and Koeppe (1996), using PET in an experimental scenario, have discovered the existence of anatomically separate areas for storage and rehearsal of verbal information in working memory. They employed a design in which participants in one condition viewed a stream of letters and were asked to judge whether a given letter matched the one they had viewed two letters back in the sequence. This task requires both storage and rehearsal. The performance of these participants was compared with that of two control groups. Participants in the first control group were asked to search for a single target letter. This ostensibly requires storage only. Participants in the second control group viewed the letter stream and were asked to make only manual responses to the presentation of a letter as they rehearsed that letter until the next one appeared. This condition requires rehearsal only.

The PET images that were obtained under these study conditions were compared with each other using a subtraction technique. In this technique, the record of brain activity obtained in one condition is subtracted from that of another. According to this logic, if the "two-back" condition calls on storage and rehearsal and the "search" condition calls on storage, the subtraction of the latter from the former should identify

the brain areas that are responsible for rehearsal. Similarly, the subtraction of the record of brain activity of the "manual" group from that of the two-back group should identify those brain areas that are involved in storage.

Using this type of analysis, the researchers found that activation of the left hemisphere's posterior parietal cortex corresponded to the storage of verbal material. Three other sites also turned up. All three were in the prefrontal cortex and corresponded to rehearsal for verbal materials. These sites were the inferior frontal gyrus (Broca's area), another region in the premotor cortex, and a region of the supplementary motor area. These areas are believed to generate a code for use in explicit speech, as we will discuss in the linguistics chapter. This code can be used to represent verbal materials for rehearsal and implicit speech as well. These two areas, the left hemisphere's posterior parietal cortex and prefrontal cortex, can be considered to form a portion of the articulatory loop that is part of Baddeley's (1986) model of working memory.

In a separate experiment, Smith et al. (1996) investigated the neural substrate of spatial working memory. Participants in this study were presented with letters that appeared at different positions around an imaginary circle. They were asked to judge whether a given letter appeared in the same position as that of a letter that had appeared three letters back in the series. The resulting PET analysis showed activation in the posterior parietal cortex, but this time in the right hemisphere. The researchers identified this region as the area where spatial information in working memory is stored. They noted that these data were in accord with those of other studies that have identified the same area as the site of storage of spatial information (Jonides et al., 1993; Petrides, Alivisatos, Meyer, & Evans, 1993).

Evidence from animal studies suggests a location for the rehearsal or maintenance of a spatial code in working memory. Goldman-Rakic (1993) tested monkeys by giving them a delayed-response task. In this procedure, a monkey is shown two locations, only one of which contains a food item. There is then a forced delay of several seconds, during which a barrier visually obscures the locations. The delay forces the monkey to maintain the spatial location of the food item in working memory. The barrier is then removed, and the monkey must choose the location that holds the food item to get the item as a reward. Monkeys whose dorsolateral prefrontal cortex has been lesioned cannot perform this task—implying that this region is the site for the maintenance of spatial information in the working memory system.

Studies with both animal and human subjects reveal a third working memory system for objects. This system codes and stores visual object representations. Whereas spatial working memory would code for the location of a letter in space—for example, whether it appears to the left or to the right—visual object memory would code for a visual representation or identity of the letter itself, which would include attributes such as its shape and color. This differentiation reflects the existence of the dorsal "where/how" pathway for location and sensorimotor interaction as separate from the ventral "what" pathway for object recognition, discussed earlier in this chapter. Data from Wilson, O'Scalaidhe, and Goldman-Rakic (1993) show that in monkeys, one brain region mediates object–shape memory and another mediates object–location memory. The principal sulcus is the area corresponding to spatial location; the inferior convexity corresponds to shape location. Both are located in the frontal area. In humans, distinct brain areas lateralized

to either hemisphere carry out these functions. Occipital, parietal, and frontal sites in the right hemisphere are linked to spatial memory, whereas parietal and IT sites in the left hemisphere are linked to object memory (Smith et al., 1995; Smith & Jonides, 1994).

Evaluating the Neuroscience of Working Memory

Jonides and Smith (1997) make a few general conclusions concerning the neural architecture for working memory. They speculate that there may be a separate working memory system for each sensory modality—each having its own storage buffer and rehearsal capacity. The visual object memory system described above would service vision, another system would serve audition, yet another gustation, and so on. Recent reviews of numerous neuroimaging results have generally confirmed that prediction (D'Esposito & Postle, 2015; Thompson-Schill, 2003). Verbal and spatial working memory, however, are modality independent. They can represent and rehearse information from both visual and auditory modalities.

These investigators are also careful to point out that the various memory codes are in the service of other cognitive processes, and they give several examples (Jonides & Smith, 1997). One could use the spatial code to form a mental map of how to get somewhere and a visual code to mentally compare the shapes of two objects. What is lacking, they argue, is a more abstract conceptual code that could represent semantic items such as words. This type of code would allow for additional cognitive tasks such as language comprehension, problem solving, and deductive reasoning. There are hints in the literature of the existence of a semantic code of this nature. Future research would be needed to determine whether such a code would have its own attendant working memory structures for storage and rehearsal.

The studies of the neural basis of working memory that we have surveyed thus far have been very informative. They show us that there are, in fact, a large number of distinct working memory systems that correspond to different sensory modalities and various abstract characteristics. This work can help us reconceptualize Baddeley's (1986) model of working memory, which originally postulated the existence of verbal and visual components only. We could update this model by incorporating these additional working memory systems. You might recall that Baddeley's model also included an executive control system. This system is responsible for the control of various aspects of working memory function. Executive processes are also used in reasoning and problem solving. For this reason, we will talk about reasoning and problem solving in the section on executive function.

Neural Substrates of Long-Term Memories

We have already reviewed some aspects of long-term memory in our discussion of learning and the hippocampal system. Remember that when information first arrives at the brain, multiple areas of the cortex—each corresponding to a unique sensory modality—process it. These representations, however, are short-lived and would fade away quickly if it were not for the hippocampal system, whose job it is to integrate the disparate aspects of a memory experience and to return the consolidated memory to the rest of

cortex. So the synaptic connections between the hippocampus and sensory and motor cortices are where some of our long-term memories reside. But there are multiple forms of long-term memory. Procedural memories store procedural or skill knowledge and are demonstrated through action. Declarative memories store factual or event knowledge and are demonstrated through explanation. The two types of declarative memory are semantic memory for facts and episodic memory for events. See Chapter 5 for a more extensive treatment of these types.

Given this variety of types of long-term memory, what can we say about their neural bases? Are there different brain areas that mediate processing with respect to the different types of long-term memory? The answer is yes. Research in this area shows that they are governed by different brain regions.

Declarative memories rely on the cortex for storage and the hippocampal system for consolidation. But two separate aspects of this system mediate semantic and episodic information. This conclusion comes from an examination of case study data. Vargha-Khadem et al. (1997) reported on patients who sustained damage to only the hippocampus early in life. In these patients, there was no damage to the limbic cortex of the medial temporal lobe. The limbic cortex consists of the parahippocampal, entorhinal, and perirhinal cortices. These areas are considered part of the overall hippocampal system and mediate connections between it and the rest of cortex. These patients could not recall anything that happened to them during the course of a day. If they went to a movie or visited a friend, they would have no memory of it. They did, however, have excellent memory for facts and did relatively well in school. This suggests that the hippocampus proper is responsible for the consolidation of episodic memory and that the limbic cortex mediates semantic memory. These researchers also concluded that if both the hippocampus and the limbic cortex were destroyed, all declarative memory ability would be lost. More research is needed to confirm these conclusions.

Learning in procedural memory corresponds to a change in the neural systems that underlie the acquisition of a given task. It is the motor regions that govern skill performance and procedural knowledge, not the widespread sensory cortical areas that hold declarative information. Also, procedural memory does not depend on the hippocampus for consolidation. The **basal ganglia** play a critical role in skill learning (Graybiel, Aosaki, Flaherty, & Kimura, 1994). This is a cluster of brain areas involved in voluntary motor responses. The basal ganglia consist of the striatum—which is made up of the caudate nucleus, putamen, and globus pallidus (see Figure 6.11). Changes to the motor cortex also have been documented as taking place during skill learning (Grafton et al., 1992; Grafton, Woods, & Tyszka, 1994). In one study, monkeys trained to pick up small objects using fine finger and hand movements evidenced increased representation of those body areas in the primary motor cortex (Nudo, Milliken, Jenkins, & Merzenich, 1996).

THE NEUROSCIENCE OF EXECUTIVE FUNCTION AND PROBLEM SOLVING

Executive function refers to cognitive operations such as planning, the sequencing of behavior, cognitive control, the flexible use of information, and goal attainment. Many

of these same operations are called on in problem solving. As we saw in our examination of the cognitive approach, problem solving consists of trying to attain a final goal via the performance of a sequence of operations that leads to the attainment of individual subgoals. In this section, we discuss the neural basis of such problem-solving abilities.

The hallmark of frontal lobe damage is a diminished capacity to perform goal-directed behaviors. Patients who have sustained this type of brain damage suffer from **executive dysfunction**, a disorder characterized by a broad array of deficits. Some of these patients have difficulties initiating actions or terminating them once initiated. For example, the patients may be listless, sitting around the house all day—but once they are engaged in some action, such as brushing their teeth, they may have difficulty stopping. This kind of behavioral deficit is referred to as **psychological inertia**. These patients sometimes appear to be impelled to perform actions that are "suggested" by the environment, such as picking up a pencil and writing with it as soon as they see it. This phenomenon in which a stimulus in the environment triggers an automatic behavior is called **environmental dependency syndrome**. Needless to say, these individuals have difficulty solving even simple problems.

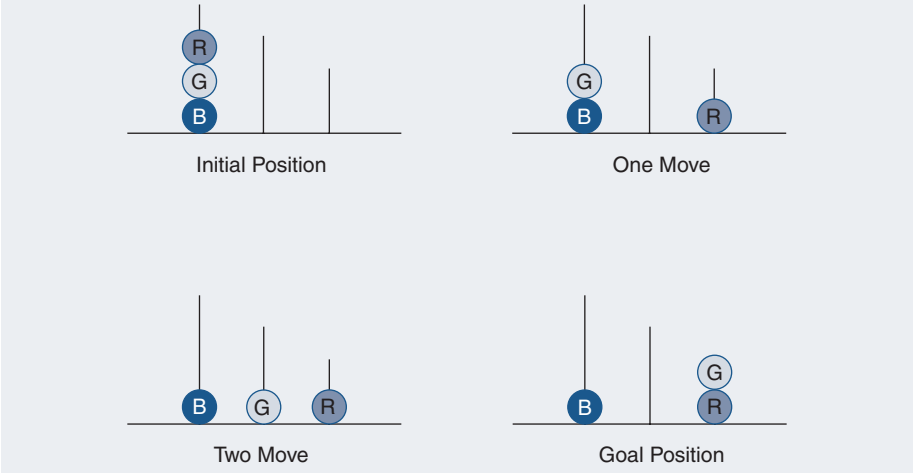
Sequencing, that is, the sequential ordering of actions, is necessary in problem solving. Any strategy that is part of problem solving includes sequencing because multiple steps are part of the problem-solving process. The Tower of London problem is a tool that researchers have used to test this capacity (Shallice, 1982). Figure 6.12 shows several Tower of London configurations. It is slightly similar to the Tower of Hanoi problem discussed in Chapter 5, but the three sticks have different heights. Three balls of different colors have holes through them and can be placed on the sticks. The leftmost stick is the tallest and can accommodate all three balls. The middle stick is intermediate in height and can accommodate two balls. The stick to the right is the shortest and can hold only a single ball. The solution to any of the Tower of London problems involves moving the balls that make up an initial starting configuration until they conform to a final target configuration, in the smallest number of moves possible.

Patients who have sustained frontal lobe damage, especially to the left frontal lobe, have great difficulty solving Tower of London problems. Shallice (1982) studied 61 patients with and without localized lesions of either frontal lobe. The patients were then given Tower of London problems. The number of problems the patients were able to solve in 1 minute was tallied. The results showed that the group with left anterior frontal lobe damage had scores that were significantly worse than those of the other groups. This area thus seems responsible for planning and sequencing of the type that is needed in Tower of London tasks.

In a neuroimaging study, healthy individuals were asked to solve a computer version of the Tower of London task. Regional cerebral blood flow (rCBF) in these individuals was measured (Morris, Ahmed, Syed, & Toone, 1993). Participants who performed the computer task showed increased levels of rCBF in the left prefrontal cortex. Those subjects who took more time to plan their moves and computed the solutions in fewer moves had even greater rCBF, suggesting that activity in this brain area is responsible for planning during problem solving.

When we examine the insights that are to be gleaned from studies of frontal lobe damage, we see a good example of convergence between cognitive science approaches.

Figure 6.12 The Tower of London task. Players must produce the goal position in as few moves as possible.



Source: Adapted from Shallice (1982).

In this case, the convergence is between neuroscience and studies of problem solving in the areas of cognition and artificial intelligence. Patients with frontal lobe damage clearly demonstrate deficits that have to do with directing themselves toward a goal or breaking down a problem into subgoals. They also have difficulty sequencing their actions appropriately. These components of problem solving are exactly what would be required in the solution of any problem and what computer programs such as General Problem Solver and SOAR do to solve problems (from Chapter 5). The fact that these ideas come up independently in different disciplines suggests that they are fundamental to an understanding of problem solving.

Theories of Executive Function

In this section, we consider two theories of executive function. The theories are cognitive in nature but have specific neurological implications. Both are formulated to account for automatic versus controlled forms of attention. **Automatic attentional processes** do not require conscious control. They occur when one is experiencing familiar situations. Automatic responses are then triggered by environmental stimuli. After years of practice, driving a car is an example of an automatic process. It can be carried out virtually without thinking or while carrying on a conversation. **Controlled attentional processes**, on the other hand, require conscious control. They are operational when one

responds to novel or difficult situations for which there is no prior learned reaction. In these instances, attention must be voluntarily directed. Problem solving generally calls for this type of processing.

Norman and Shallice (1980) proposed a model that relies on **action schemas**, structures that control automatic attentional processes. In this model, a schema can be triggered either by perceptual input or by the output of other schemas. An example of the former would be the image of your unmade bed in the morning. The sight of it could activate a bed-making schema. An action schema can be thought of as an if–then production rule where perceptual inputs serve as the preconditions necessary to trigger a context-specific action.

In a more general sense, a schema can be thought of as a stored framework or body of knowledge on the subject of some topic (Schank & Abelson, 1977). This type of rule-based schema is sometimes referred to as a **script**. The most often used example is a script for eating in a restaurant. It would consist of a sequence of events: being seated, getting menus, ordering, waiting, eating, paying the bill, and so on. Scripts can be used for solving problems because they specify goals and the sequences of actions needed to achieve them. At first glance, it may be difficult to imagine precisely how explicit rules and scripts are encoded in brain tissue. However, computer simulations of neural networks, modeling the interactions between prefrontal cortex and subcortical regions, have demonstrated that approximations of rule-like behavior can emerge from neural systems (Noelle, 2012; O'Reilly, 2006). For an in-depth exploration of neural network simulations, see Chapter 7.

In the Norman and Shallice (1980) model, schemas can be activated independently of one another. This can pose a problem, since in everyday situations, usually just one action needs to be performed at any given time. The model specifies two ways in which a subset of appropriate schemas can become activated. The first is **contention scheduling** and is used to govern routine habitual performances as well as the performance of novel tasks. The second selection mechanism is the **supervisory attentional system (SAS)** and is used only for nonroutine actions.

In contention scheduling, each action schema is executed when the level of activation in one of its control units reaches a certain threshold. However, each schema has mutually inhibitory connections to others so that the more one schema is activated, the more it suppresses the activation of those it is connected to. This prevents the simultaneous activation of multiple schemas and ensures that the most appropriate one—the one receiving the greatest amount of activation—is triggered. Contention scheduling ensures that you don't try to do two things at once, such as trying to step on both the accelerator and the brake while driving. An important component of executive function is cognitive control, which involves selecting preferred actions and inhibiting disruptive actions. Neuroimaging research by Kan and Thompson-Schill (2004) suggests that the left prefrontal cortex plays a major role in this process of selecting among competing percepts, competing concepts, and therefore competing actions.

In the solving of problems that are related to new situations, existing schemas can fail or there may be no learned schema that can be applied. In these cases, the contention system needs to be supplemented with an alternate selection process. According to Norman and Shallice's (1980) model, this occurs as the additional activation of other

schemas from the SAS. The SAS contains general programming or planning systems that are slower and more flexible than the fast, automatic, and unchanging schemas of contention scheduling. These schemas are designed to be more general purpose and applicable to a wide variety of problem types.

Joaquin Fuster (2015) warns against treating the executive functions of the prefrontal cortex as though they are the *original source* of goals and plans, as that would falsely assume that those goals and plans have no preceding causes. Every event has some form of causal precursor. Fuster describes this attentional component of executive function as best understood as an "enabler" of causal influences in the perception–action cycle, not a generator of those causal influences (see also Hazy, Frank, & O'Reilly, 2007). It would be a mistake to think of the executive functions of prefrontal cortex being like the CEO of a company, receiving reports from and delivering orders to all the other brain regions. Instead, the prefrontal cortex is like an airline hub airport where most of an airline's flights tend to travel through as they go from one place to another. The prefrontal cortex and its executive functions play a special role in coordinating the transmission of signals between sensory and motor brain areas, but it is not quite *in charge of them*. In that light, one can think of Norman and Shallice's SAS as a higher-level monitoring system, one that can apply general strategies to bear on a problem that the perception–action system is struggling with. The SAS can also suppress or turn off inappropriate schemas that might be triggered inadvertently. The failure to suppress these inappropriate schemas can result in "capture errors." Reason (1979) describes a person who, while walking across his back porch on the way to his car, began putting on his boots and jacket for working in the garden. Perhaps you can recall a similar situation from your own personal experience. Momentary lapses of attention seem to correspond to reduced activity or monitoring in the SAS.

So where in the brain are the different components of this model located? The left prefrontal cortex is the probable location for the SAS. As mentioned above, lesions in this region are associated with difficulties in solving Tower of London problems. Damage to this region would explain the pattern of symptoms we see in executive dysfunction, where patients have trouble with problem solving but little difficulty performing well-learned tasks. This would also explain environmental dependency syndrome. Here, the contention scheduling system is clearly operating, as environmental stimuli automatically trigger their learned responses. These responses are out of place, however, as they cannot be overridden by the supervisory system. It could also be that schemas or some aspects of schematic knowledge may be stored in the frontal lobes. This makes sense, as disruptions to the frontal areas result in the failure to apply and use schemas and in a consequent deficit in problem-solving ability. Further discussion of central executive functions, such as reasoning and decision making, also takes place in Chapter 8, where we discuss the evolutionary approach.

Stuss and Benson (1986) offer a second theory of executive function involving controlled and automatic attention. In their view, there is a three-tiered hierarchy of attentional systems. The automatic system that corresponds to the lowest level makes associations between sensory representations and is governed by posterior brain areas. The supervisory system corresponds to the middle level and runs executive processes and is used in problem solving. It resides in the frontal lobe. In addition, Stuss and Benson

postulate a metacognitive system that corresponds to a third level. **Metacognition** refers to any process that monitors, regulates, or controls any aspect of cognition. Metacognitive regulation includes planning, resource allocation, checking, and error detection and correction (Brown, Bransford, Ferrara, & Campione, 1983). The prefrontal cortex is believed to be the site of the metacognitive system. In a problem-solving situation, metacognitive processing would evaluate whether a particular strategy is working and, if it decided that it was not working, would initiate consideration of another strategy. Individuals lacking metacognitive control would persist in their applications of inappropriate strategies, meaning that they would get stuck on one approach to a problem and fail to consider other options.

OVERALL EVALUATION OF THE NEUROSCIENCE APPROACH

Neuroscience provides us with a straightforward description of the brain that includes microscopic-level descriptions of individual neurons and larger-scale descriptions of brain structures. Although we have not mentioned it, comparative neuroscience compares these structures across species to see how evolutionary pressures have changed them. In this sense, neuroscience and the evolutionary approach go very well together. For more on this, go to the section on comparative cognition in the evolutionary approach chapter.

Of course, physical structure cannot be divorced from function, so neuroscience also addresses physiology. But physiology is best described within some organizing framework. In recent years, that framework has become cognitive processes, so we see the birth of cognitive neuroscience that attempts to map cognitive function onto brain structure and function. Just as with the cognitive approach, the neuroscience researchers have produced various models. These neuroscience models differ from their counterparts in that they specify the locations in the brain where information is represented and processed. They also specify the pathways by which information is passed from one processing center to another. The integration of the two types of models is a fruitful procedure for future investigations.

SUMMING UP: A REVIEW OF CHAPTER 6

1. Neuroscientists study nervous system structure and function. Researchers in the field of cognitive neuroscience attempt to relate these findings to cognitive processes.
2. A number of different methods are used in neuroscience. Case studies of human patients and lesion studies with animals investigate the behavioral deficits that follow damage

to neural tissue. Traditional brain recording techniques include single- and multiple-unit recording, in which the activity of single or multiple neurons is recorded. The EEG yields a global pattern of brain action. When these are taken in response to a stimulus, they are called event-related potentials. More modern brain imaging methods include PET, fMRI, MEG, and the KESM. Electrical and magnetic stimulation of brain tissue are also possible. One of the newest forms of brain stimulation is called optogenetics, which allows researchers to directly stimulate neurons using light pulses.

3. A neuron can be considered the functional unit of the brain, as it receives messages from other neurons and can also send messages. These messages are in the form of an action potential—an electrical signal that starts at the cell body, travels down the axon, and ends at the terminal button. The action potential triggers the release of neurotransmitter molecules that diffuse across the synaptic cleft, attach to receptors on the membrane surface of a receiving cell, and initiate postsynaptic potentials. These potentials are summated at the axon hillock and regulate the neuron's firing rate.

4. The cortex consists of four lobes. The frontal lobes govern motor behavior and problem solving. The temporal lobes are involved with hearing and speech comprehension. The parietal lobes underlie somatosensory perception and attention, while the occipital lobes are responsible for initial processing of visual information.

5. There are two visual system–processing streams. (1) The dorsal pathway handles information regarding location and motion. (2) The ventral pathway is concerned with object recognition.

6. Visual deficits, or agnosias, result from brain damage to cortical recognition areas. In apperceptive agnosia, patients cannot assemble features together to create a whole object percept. In associative agnosia, patients can perceive an entire object but have difficulty naming or assigning a verbal label to it. Prosopagnosics cannot identify faces and suffer damage to the fusiform gyrus in the temporal lobe.

7. A number of different brain structures underlie attentional processing. These include the RAS, the superior colliculus, the thalamus, the intraparietal sulcus, the cingulate cortex, and the frontal lobes. In the component process model, these areas are functionally specialized but work together to govern attention. In the distributed network models, there is some functional overlap between the areas.

8. Brain damage to these regions can produce attention disorders. In hemispatial neglect, patients have difficulty attending to the left side of objects because of damage to the right superior temporal gyrus. ADHD does not have clear-cut physical causes but is heritable. Symptoms include difficulty focusing, impulsiveness, short temper, and difficulties in following plans.

9. The engram is the physical location of a memory. In animal lesion studies, it has been difficult to locate the engram. Greater tissue damage is associated with greater memory loss, but there does not appear to be a single place in the brain where memories are deposited. Memory seems to be laid down in many different parts of the brain, a principle called equipotentiality.

10. Memory circuits in the brain are created by repeated high-frequency stimulation of synapses. This LTP strengthens these synapses so that they are more easily activated in the future.

11. The hippocampus is the site where information gets transferred from working

to long-term memory, a process called consolidation. Damage to the hippocampus can result in anterograde amnesia, an inability to consolidate new information.

12. Brain imaging studies show that there are multiple sites in the brain for storage and rehearsal of visual and auditory information.
13. Declarative memories are stored in cortical regions. Episodic memory, in particular, seems to be consolidated by the hippocampus, whereas the limbic cortex mediates semantic memory. The basal ganglia are implicated in procedural knowledge and skill acquisition.
14. Executive dysfunction is a product of frontal lobe damage and is characterized by problems in initiating and terminating actions.
15. Research suggests that there are two separate executive systems. Action schemas are hypothesized to control automatic attention, while the SAS governs the performance of novel tasks that have not yet been automatized.

SUGGESTED READINGS

Banich, M. T. (2004). *Cognitive neuroscience and neuropsychology.* Boston, MA: Houghton Mifflin.

Boleyn-Fitzgerald, M. (2010). *Pictures of the mind: What the new neuroscience tells us about who we are.* Upper Saddle River, NJ: FT Press.

Eichenbaum, H. (2002). *The cognitive neuroscience of memory.* New York, NY: Oxford University Press.

Goldberg, E. (2002). *The executive brain: Frontal lobes and the civilized mind.* New York, NY: Oxford University Press.

Posner, M. I. (2004). *Cognitive neuroscience of attention.* New York, NY: Guilford Press.

Quiroga, R. Q. (2017). *The forgetting machine: Memory, perception, and the Jennifer Aniston neuron.* Dallas, TX: BenBella Books.

CHAPTER SEVEN

THE NETWORK APPROACH

Mind as a Web

THE NETWORK PERSPECTIVE

The network approach is influenced by the principles of operation and organization of real-world brains. All biological brains are made up of cells called neurons that are wired to one another in a complicated fashion. Activity in one neuron or a set of neurons in turn activates other neurons through these connections. It is this activity that underlies all mental operations—whether it be recognizing your cousin's face, calculating the tip on a restaurant bill, or deciding which law school to attend. **Connectionism** is a field of study in the network approach. Connectionists try to understand how the mind performs these kinds of operations via the construction of an **artificial neural network (ANN)**, which is a computer simulation of how populations of actual neurons perform tasks. Semantic and propositional networks constitute another field of study under the network approach. They look at frameworks for how concepts and sentence-like representations might be implemented in networks. At the end of the chapter, we will examine an emerging discipline called network science that analyzes the structure and function of all networks.

The use of ANNs brings up a host of interesting issues in cognitive science. Perhaps the most interesting of these is the issue of knowledge representation. The dominant paradigm in cognitive science is the use of symbols to represent information. These symbols are then represented, stored, and operated on by cognitive or computer processes. In the case of ANNs, information is represented not in the form of a symbol but as a pattern of activation in the network. The classical symbolic representational view and the connectionist view also differ in their functional architecture. Traditional notions in cognitive psychology and machine intelligence have processing occurring in serial stages, whereby information is transferred from one relatively large-scale system and fed to the next in the sequence. In the network

Learning Objectives

After reading this chapter, you will be able to:

1. Compare serial to parallel processing in computing systems.
2. Calculate how artificial neural network nodes send activation to one another.
3. Contrast supervised (backpropagation) and unsupervised (Hebbian) learning algorithms in neural networks.
4. Describe the structure and function of Elman's simple recurrent network.
5. Explain McClelland and Rogers's network simulation of hierarchical category knowledge.
6. Compare interaction-dominant dynamics and component-dominant dynamics.
7. Distinguish random networks from small-world networks.

approach, by contrast, processing events are carried out in parallel by many small processing units, and they transfer information back-and-forth in multiple directions.

In the second part of this chapter, we shift from a focus on information *processing* (or the flow of information in a neural network) to information *representation* (or the storage of knowledge in a semantic network). This gives us an opportunity to compare neural networks and semantic networks. To date, most ANNs are capable of somewhat limited representation—enough to enable the implementation of tasks such as pattern classification and recognition of temporal structure. These limited representations may not adequately reflect the full complexity of human cognition. To accommodate this limitation, a different kind of network may be needed, one that is capable of storing and combining knowledge in the broadest sense. Using a more abstract format of representation that does not address how neural systems work, semantic networks model how information in a permanent memory store, such as human long-term memory, might be structured. They do this through their use of a rich set of hierarchically structured concept and concept-property nodes to represent knowledge.

Whereas ANNs are processing dominant, semantic networks are representation dominant. Researchers in the first discipline are concerned with how to get a system to do something such as recognize a face, learn a language, or solve a math problem—while adhering to some degree of biological plausibility. Connectionists aim to build models the provide insight into how groups of neurons produce cognition. The primary goal in this field is to achieve results: to take some information as input, perform a computational process on it, and produce an output that is judged as being correct. The focus is on functionality. Researchers in the second field are concerned with how information is ordered, how symbols and other types of representations are organized with respect to each other. The focus here is on abstract structure, in particular, how information is stored in our long-term memories and how it is interrelated—with less concern for adhering to neural plausibility.

The third and final part of this chapter will look at an emerging paradigm in the sciences, one that is called network science. This field of study shows that networks are not just limited to the traditional fields of cognitive science. Surprisingly, very different networks have been found to act alike. It seems crazy, but the routes airlines fly, the ways by which diseases spread, and the way in which voice data are transmitted on cellular telephone systems all share similarities with the way information is processed in the brain. In this last section, we examine the way other networks operate and show that these can inform human mental processes like the way the visual system assembles features or how we "brainstorm" and think creatively. Network science also gives us new insights into classic issues in cognitive science, such as whether we have a single "self" (i.e., a unified center of mental control) or are in fact made up of a collection of competing control centers, a topic we took up in the philosophy chapter.

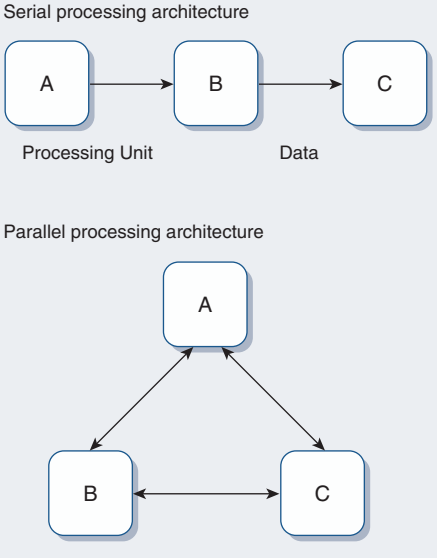
ARTIFICIAL NEURAL NETWORKS

Traditional computers are **serial processors**. They perform one computation at a time. The output of a particular computing unit can then serve as the input to a second

computation, whose new result serves as the starting point for yet another computation, and so on (see Figure 7.1). The brain, as well as ANNs, relies on a completely different processing strategy. Here, large numbers of computing units perform their calculations in parallel. One computing unit does not need to wait for another to finish its computation before it can begin its work. As shown in Figure 7.1, these units that operate in parallel are also not limited to receiving inputs from only a single unit: They can receive and process multiple inputs and transmit multiple outputs. This type of architecture is referred to as **parallel distributed processing (PDP)**.

Researchers in the field of artificial intelligence and, more generally, in computer science solve a particular problem by constructing an algorithm, or a procedure for solving it. Every detail and step of the procedure is planned or anticipated ahead of time. One conceptualizes the problem and its solution in terms of symbolic representations and transformations of these symbolic representations. This is referred to as the

Figure 7.1 **The top panel represents a serial processing architecture. Boxes are processing units, and arrows are data. Each box performs a computation on the input it receives. The result of the computation is transmitted as an output to another processing unit. Unit B cannot start its computation until it receives its input from Unit A. The bottom panel shows a parallel processing architecture. Each processing unit can be connected to every other unit. Computation in the units occurs simultaneously (i.e., in parallel).**



knowledge-based approach to problem solving, and it is used in many artificial intelligence approaches. In contrast, researchers using ANNs are more concerned with the overall behavior of a network. They leave the computational details to the network itself and don't pay much attention to symbolic representations or rules. This is known as the **behavior-based approach** to problem solving.

ANNs are pretty good at solving certain kinds of problems. Among them are problems of classification. These involve producing a learned label for a stimulus (pattern recognition) and assigning stimuli to categories (a form of concept formation). ANNs are also fairly adept at solving control problems—for example, the programming of the movements of a robot arm—and problems of constraint satisfaction—for example, the devising of airline schedules. In these situations, the data are often "noisy," the problem is usually ill defined, and the outcome may be unclear. There are, though, many scenarios in which the problem is well defined and the solution process is clearly understood. In these scenarios, a conventional rule-based system, such as the ones used in traditional artificial intelligence, may be preferred.

But let's get back to the issue of symbols. Why is it apparently okay for connectionists to ignore them? After all, it seems impossible that any kind of problem could be solved without concepts being represented in some fashion. The answer is that representations are inherent in ANNs but do not exist in them in the form of symbols. They exist in most networks as a pattern of activation among the network's elements. This is known as a **distributed representation**. To illustrate, the idea of "banana" in a connectionist model might be represented by the simultaneous activity of the different elements—or Nodes A, B, and C—and the connections among them. That said, some ANNs do represent concepts via activity in single nodes to make analysis of their activity easier. In these networks, activity in Node A alone would represent, let's say, "apple." This is a form of **localist representation**, where the connectionist assumes that the single node is a "stand-in" for a population of neurons. In either case, though, we need not employ the idea of a symbol, such as a letter, to stand for the concept. The concept is instead represented as the activity of the nodes or node and the interactivity between nodes (e.g., the bottom half of Figure 7.1) deviates substantially from a serial symbolic processing system (e.g., the top half of Figure 7.1). We have already discussed both types of these representations and how they might code for visual objects in the neuroscience chapter, where we referred to them as distributed coding and localist coding.

One advantage of neural nets is that they are capable of learning. That is, they can adaptively change their responses over time as they are presented with new information. However, it should be noted that ANNs are not the only form of simulated cognition capable of learning. Learning is also evidenced in machines that use symbolic methods. The acquisition of new information or a new skill usually requires repetition. For example, a child learning the multiplication tables will repeat them over and over again to himself or herself. A child who makes a mistake must correct himself or herself and then repeat the new information. Neural networks operate on these same principles. They learn over a series of trials to perform a task or to come up with an answer to a question. If they produce a wrong answer, the correct answer can be "shown" to them. This feedback is then used to adjust the performance of the network until it produces the correct answer.

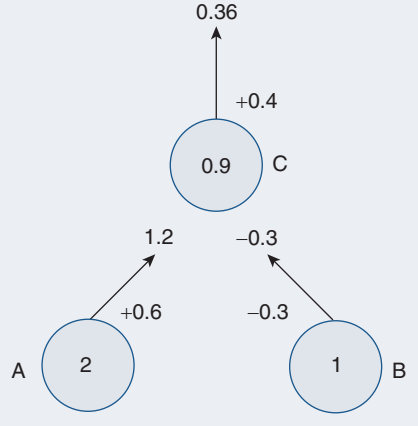
CHARACTERISTICS OF ARTIFICIAL NEURAL NETWORKS

Real neural networks exist in the brain in the form of neurons and the connections between them. The artificial networks constructed by connectionists typically exist only as software simulations that are run on a computer. (However, some ANNs are physically implemented on computer chips, e.g., Merolla et al., 2014.) Each neuron, or basic computing unit in an artificial network, is represented as a **node**, and the connections between nodes are represented as **links**. Each node has an activation function that determines whether (or how much) it is active as a result of its incoming signals. A node, if it is stimulated, sends out a signal, represented as its activation value multiplied by the strengths (positive or negative weights) of the synaptic links it has with other nodes.

Links in a neural network have **weights**, which specify the strength of the link. A weight can be positive, negative, or zero. A positive weight on a link functions like an excitatory synapse and a negative weight functions like an inhibitory synapse. So, for example, a unit with an activation value of 1 passing its information along a link that has a weight of 0.5 will have a net output of 0.5. It will positively stimulate the node to which it is connected by a factor of 0.5.

A unit with an activation value of 1 passing along a link that has a weight of –0.5 would have a net output of –0.5, which would inhibit whatever node that link connects to. Most nodes receive many outputs from other nodes. They sum these positive and negative inputs to determine their activation level. Figure 7.2 shows a simple neural

Figure 7.2 **A simple neural network. Activation values are shown inside each node. Weights are indicated on the links and net outputs at the endpoint of each connection. Node A has an activation value of 2 and a weight of 0.6. Its net output is thus 1.2. Node B's activation value is 1 and its weight is –0.3, yielding a net output of –0.3. These outputs are summated at Node C so that it has an activation value of +0.9. The activation value of Node C multiplied by its weight of 0.4 produces an output of +0.36.**



network with activation values, weights, and net outputs. Node C in Figure 7.2 uses a linear activation function, but many ANNs use a nonlinear activation function.

EARLY CONCEPTIONS OF NEURAL NETWORKS

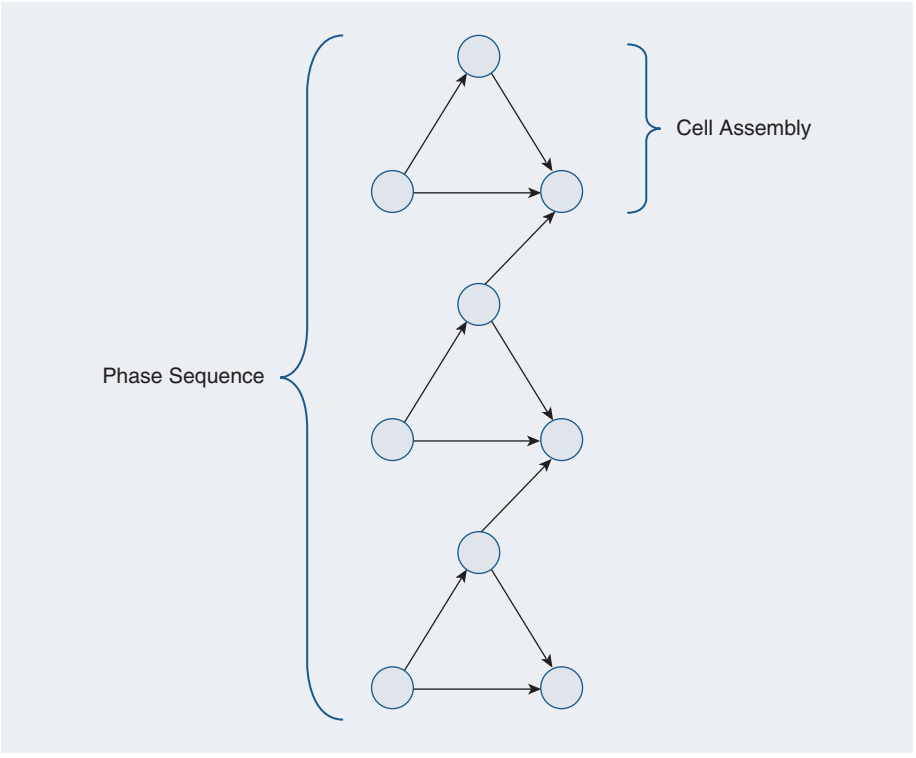
In 1943, Warren McCulloch and Walter Pitts were the first researchers to propose how biological networks might function. They made a few simple assumptions about how neurons might operate. They assumed that each neuron had a binary output—that is, it could either send out a signal, corresponding to its being "on," or not send out a signal, corresponding to its being in an "off" state. Whether or not a neuron would fire was determined by a threshold value, thus making this a nonlinear activation function. The weights of the connections between neurons, in their model, were additionally assumed to be at a fixed value. Networks operating under these rules are capable of computing simple logical operations, such as *or*, *and*, and *not*. Digital computers of the sort that sit on our desks perform calculations that are based on these simple logical functions.

Donald O. Hebb (1949) was the first person to propose how changes among neurononal connection strengths (i.e., weights) might explain learning. According to the Hebb rule, when one cell repeatedly activates another, the strength of the connection between the two cells is increased. Neurons that fire together wire together. In this fashion, the connectivity pattern among neurons changes to include newly acquired knowledge. These circuits are believed to be the neural foundation of learning and memory. Imagine being at a party and trying to remember someone's name. How would you do it? Imagine that at the party, you are trying to remember a phone number you had been given. If you could not write the number down, you would repeat it over and over again—a process called rehearsal. Each repetition would correspond to a new activation of the circuit, strengthening it further. The circuit itself, once strong enough, would represent the telephone number. Retrieval of the number at a later date would then be equivalent to a reactivation of the circuit. This kind of learning algorithm in a neural network relies on correlations among firing neurons and does not require an error signal or "teacher." It is called an **unsupervised learning algorithm**.

Hebb (1949) defined two types of cell groupings. A **cell assembly** (or "population code") is a small group of neurons that repeatedly stimulate one another. A **phase sequence** is a group of connected cell assemblies that fire synchronously or nearly synchronously (see Figure 7.3). If a cell assembly coded for simple perceptual qualities, such as "red" or "round," then these qualities could become linked so as to form a phase sequence during learning and code for a higher-order concept, such as "apple."

Beginning in the 1950s, research on neural nets focused less on logical operations and more on mimicking real biological functions. An artificial nervous system called the **perceptron** was introduced (Rosenblatt, 1958). Perceptrons are neural nets designed to detect and recognize patterned information about the world, store this information, and use it in some fashion. Perceptrons are also characterized by their ability to learn from experience: They can modify their connection strengths by comparing their actual output with a desired output called **the teacher**. Thus, unlike Hebbian networks, perceptrons use a **supervised learning algorithm**. The ANN networks discussed in subsequent parts of this chapter constitute expanded versions of perceptrons.

Figure 7.3 A phase sequence consisting of cell assemblies.

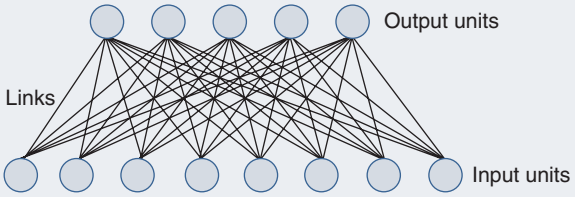


The earliest perceptron was an artificial retina called the "Mark I" (Rosenblatt, 1958). This network could recognize simple visual patterns, such as vertical and horizontal lines. It was capable, with training, of producing a desired output for each different type of visual pattern. The first perceptrons were quite simple. Each contained a single layer of weights connecting an input layer to an output layer (see Figure 7.4). The limitations of these fledgling perceptrons soon became clear. A major flaw was their inability to learn certain nonlinear logical functions (Minsky & Papert, 1969). This was in part due to their relatively weak computing power. A single layer of connection weights does not provide for much in the way of complexity and flexibility. The solution, of course, was to build more complicated networks.

BACKPROPAGATION

In a multilayer perceptron, the computing units or nodes are organized into three distinct groups. A representation of the stimulus is presented to the **input layer**. These units send signals to a **hidden layer**, which in turn feeds activation to an **output layer**. The output layer generates a representation of the response. Figure 7.5 depicts a

Figure 7.4 An example of an early perceptron with two layers. Notice that each input unit maps onto every output unit.



multilayer perceptron. A detailed description of a multilayer perceptron that learns through a nonlinear supervised learning algorithm called **backpropagation** follows.

Here's how a multilayer network designed to recognize letters works: The capital letter *A* activates the nodes in the input layer. These nodes send activation via connection weights to the hidden layer. The nodes in the hidden layer send signals via their weights to nodes in the output layer. The pattern of activation in the output layer is the network's initial response to the letter. This response is then compared with the target response, represented by the teacher. The difference between the actual output and the desired output, the **error signal**, is then used to modify the connection weights. Figure 7.6 shows these different steps that are part of the training of a multilayer perceptron. The modified weights allow the network (the next time it "sees" the letter *A*) to generate a response that is closer to the desired one. After repeated presentations of the stimulus in the presence of feedback, the network is able to produce the target response. It has, in effect, learned to recognize the letter. This kind of training based on error feedback is called the **generalized delta rule** or the backpropagation learning model.

Figure 7.5 A three-layered neural network with input, hidden, and output units.

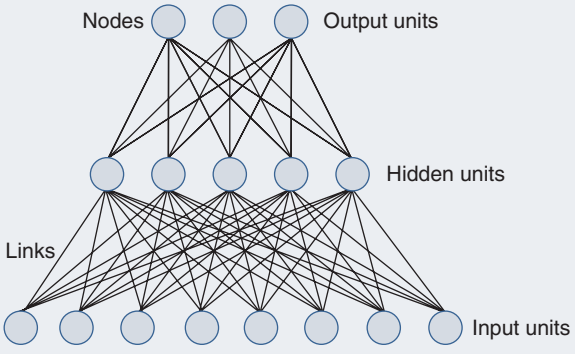
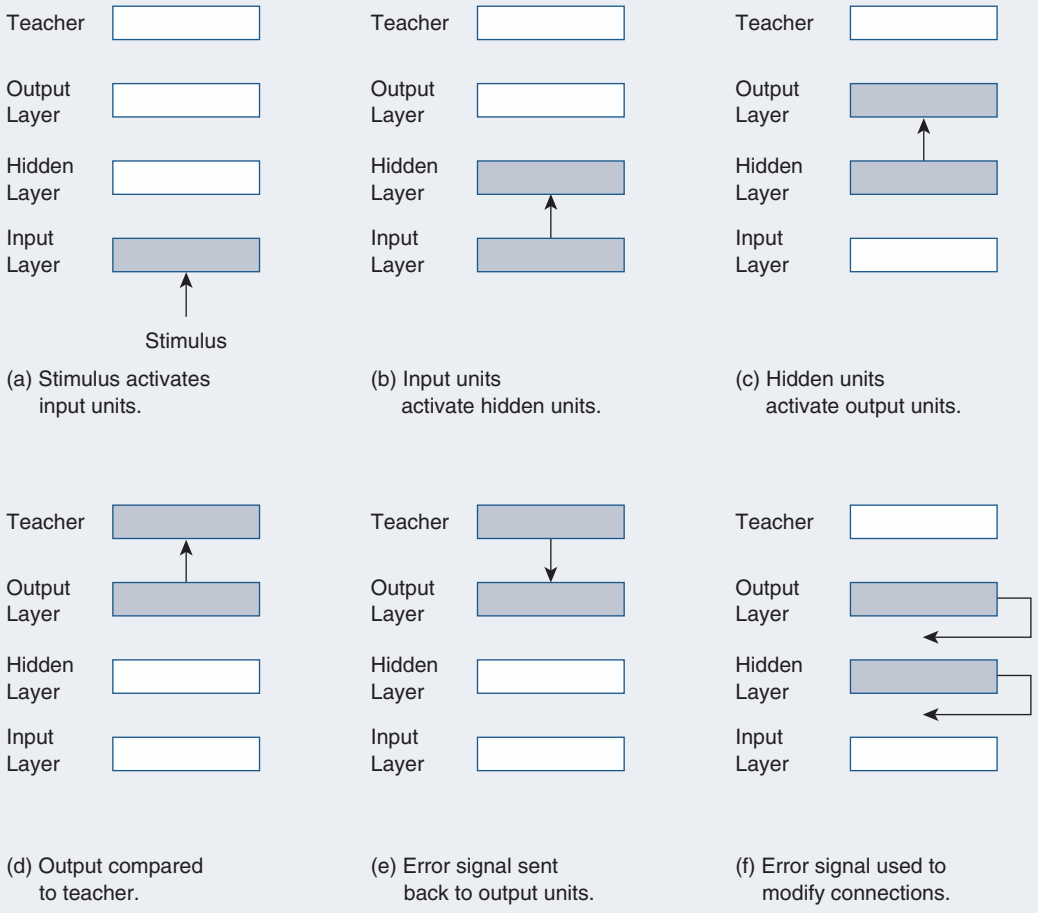


Figure 7.6 Steps in the training of a three-layered network using the backpropagation learning model.



NETtalk: An Example of a Backpropagation Artificial Neural Network

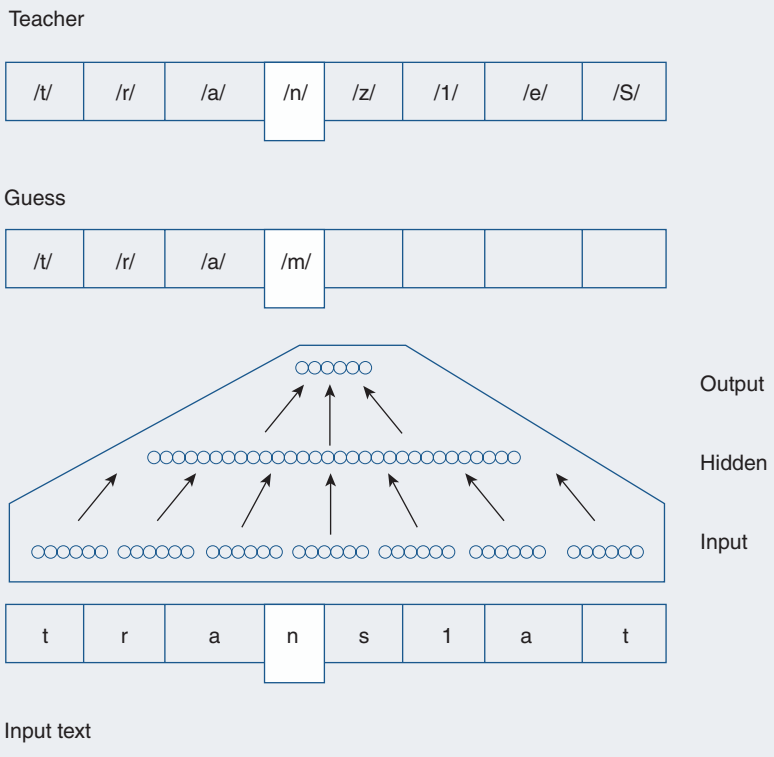
NETtalk is an ANN designed to transcribe written English into spoken English (Sejnowski & Rosenberg, 1986). It is presented with written letters of the alphabet. Its output is the correct pronunciation of the sounds represented by the letters, which is then fed to a speech synthesizer for the production of the sounds. Unlike artificial intelligence programs that existed at the time it was developed, NETtalk actually learned to make correct pronunciations after it had been supplied a given number of examples.

Although the network could not understand the meaning of what it was reading, it serves as a good demonstration of what networks of its kind are capable of and how they can serve as models of human learning.

The NETtalk system consists of three layers of nodes. The input layer has seven groups, each containing 29 individual units. The overall activity of these 29 units specifies one letter. NETtalk thus processes seven letters at a time. It focuses, though, on the fourth and middle letters of these seven. It is these target letters that the network attempts to pronounce. The other surrounding letters serve as a context and to help disambiguate the correct pronunciation, due to the fact that in English, the sound of any given letter depends heavily on the adjacent letters. Figure 7.7 shows a diagrammatic representation of the network.

The input units next connect to a hidden layer of 80 nodes, which carry out a partial recoding of the input data. These in turn connect to 26 nodes in the output layer, whose pattern of activation represents the system's initial response to the letters. This response

Figure 7.7 A diagrammatic representation of the NETtalk network, capable of converting written English text into speech.



Source: Adapted from Sejnowski and Rosenberg (1986).

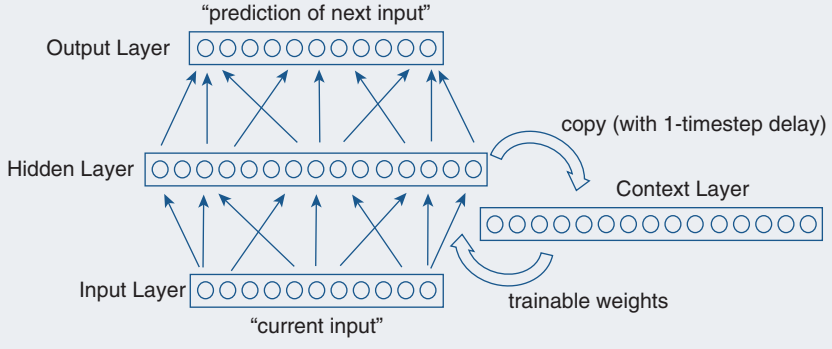
is, in effect, the network's first guess at pronouncing the target letter. The guessed pronunciation is then compared with the correct response, specified by a teacher. Weights are then adjusted using the backpropagation algorithm. Finally, the "window" or read head that viewed the seven letters at a time would move one letter to the right, to process the next letter in the text. The total number of weighted links that directly connect one node to another in NETtalk is 18,829.

How does NETtalk perform? After 50 passes through a sample text, which corresponded to about 250,000 letter-to-pronunciation pairings, it achieved a 95% accuracy rate. When the network was then tested on a new sample text without further training, it performed at 78% accuracy. This relatively impressive performance on the new text demonstrates the network's capacity for generalization; the network had extended its abilities to new words and letter combinations. NETtalk was also found to perform well after simulated damage. The simulated damage was in the form of changes made to the weights: Every weight was changed to a new random value that was within some prespecified range of the old value. When the randomized values were less than 65% of the average, performance was hardly affected—showing that the network is resistant to degradation.

The Elman Net: An Example of a Simple Recurrent Network

The Elman net was developed by Jeff Elman (1990, 1991) as a backpropagation network that uses an internal feedback loop to provide some temporal context for how best to process the current input. Instead of using a temporal "window" on the input layer, as NETtalk did, the Elman net takes each individual time slice of signal by itself on its input layer. It then incorporates temporal context by using a feedback loop (with a delay) at the hidden layer to combine the current input with its most recent internal state (see Figure 7.8).

Figure 7.8 A diagrammatic representation of an Elman net. This simple recurrent network uses backpropagation to train the weights between layers. By using temporal context to help it predict the next input, it learns the structures embedded in the temporal sequences to which it is exposed.



A key advancement of the Elman net is that, while it uses the supervised learning algorithm backpropagation, it does not actually need an explicit supervisory teaching signal. Since it learns by comparing its prediction of the next input to the actual next input, all it needs to do is wait for that next input to happen, and the difference between "actual" and "predicted" serves as the error signal for changing connection weights. Thus, rather than a "teacher" telling the output layer what the desired output activation pattern was supposed to be, the Elman net can essentially be "eavesdropping" on a sequence of inputs without the knowledge of the system generating those inputs. The Elman net generates its own error signal without the need for a teacher.

Elman nets have been used to learn the rules of letter sequences in words and the grammar of word sequences in sentences (Elman, 1990), as well as the relationships between word meanings and grammar (Howell, Jankowicz, & Becker, 2005). Thus, they learn important aspects of language structure without needing an explicit teacher. Elman nets have helped inspire a great deal of research that uses predictive coding (instead of explicit teaching) as a method for learning in a wide variety of perceptual and cognitive processes (Clark, 2015; Friston, 2019).

EVALUATING THE CONNECTIONIST APPROACH

Advantages

The connectionist approach has many strengths. The most important one is the similarity between network models and real-life neural networks—what is termed **biological plausibility** (Stillings et al., 1995). This plausibility can be demonstrated in three fundamental ways. First, artificial networks share general structural and functional correlates with biological networks. Second, and as pointed out previously, artificial networks, like their biological cousins, are capable of learning. Third, artificial networks react to damage in the same way that human brains do. In this section, these advantages and other advantages of networks are discussed in depth.

Many of the elements of artificial networks have their counterparts in biological networks. The nodes are roughly equivalent to neurons, whereas the links are analogous to axonal and dendritic connections. The functioning of constructed neural networks is also very biological. At a microscopic scale, the nonlinear activation functions that specify when and how an individual node fires, or what its average firing rate is, are inspired by actual neuron physiology. At a macroscopic scale, the parallel, distributed processing of connectionist networks is also found in the brain.

Learning in connectionist models takes place via the adjustment of the weights of the links between nodes. Work in neurophysiology shows that this process is not far off from what happens in real brains. Recall from the preceding chapter our discussion of synaptic plasticity and how it is mediated by long-term potentiation. If two cells in a circuit are simultaneously active, the synapse between them is strengthened. This increase in synaptic strength corresponds to a numerical increase in the weight of a connection between nodes in an ANN.

A third argument in support of the biological plausibility of connectionist models is the way they respond to damage. When they rely on distributed coding, neural networks demonstrate **graceful degradation**. This is a gradual decrease in performance with increased damage to the network. (By contrast, in a traditional artificial intelligence rule-and-symbol program, if a single rule is removed, the program may not work at all anymore.) In graceful degradation, small amounts of damage engender only small reductions in performance, whereas greater damage produces correspondingly larger deficits. Human patients who have sustained brain damage often show evidence of the same phenomenon. Graceful degradation is a property of a parallel computing architecture. If some nodes are destroyed, others can "take over." In other words, the network can adapt to the damage by calling on its distributed function across the existing nodes and links. Of course, this advantage holds only for nets with distributed representations. The simplified networks that use localist representations are more subject to massive failure in the event of damage.

Connectionist networks display three other interesting "psychological" phenomena: (1) pattern completion, (2) interference, and (3) generalization (Green & Vervaeke, 1996). **Pattern completion** refers to a distributed representation's ability to "fill in the blanks" when presented with incomplete sensory input (recall Figure 4.4 in Chapter 4). When a population of neurons codes for a particular stimulus, but only part of that stimulus is presented, only a portion of the population code (or "cell assembly") is made active. Then, the connections between those neurons will often cause the inactive neurons to become active, thus fulfilling the whole population code and "filling in the blanks" of the percept. **Interference** refers to instances in which two sets of information that are similar in content interfere with each other. If a student were studying Spanish and Italian at the same time, he or she might forget words in one language that are similar in sound to words in the other language. In a related fashion, neural networks trained to learn large numbers of patterns show signs of interference: They have difficulty distinguishing similar patterns. **Generalization** refers to the ability to apply a learned rule to a novel situation. If one had learned that (almost) any animal that had wings was a bird and then saw a new animal for the first time that also had wings and called it a bird, he or she would be generalizing. Rumelhart, McClelland, and PDP Research Group (1986) trained a network to learn the past tenses of verbs. The network was even able to conjugate new verbs that it had never seen before. For instance, it was able to produce the past tense *wept* for *weep* because it had been trained to associate *slept* with *sleep*. The network was capable of bona fide generalization.

Problems and Disadvantages

We have looked at the many advantages of the connectionist approach, but it is also limited, in a number of respects. What at first seems like its primary strength, biological plausibility, should also be viewed as "a work in progress." If we make a more detailed comparison of ANNs and actual brains, we discover a number of significant differences. First, real neurons have far more complex activation functions than what most simulated neural networks use. For instance, the relative timing of incoming synaptic signals (on

a millisecond timescale) can substantially alter the likelihood of a neuron producing an action potential and the likelihood of it benefiting from the Hebbian learning process (Caporale & Dan, 2008). Moreover, when different neurotransmitters wash over a biological neural network, they can temporarily alter the connection weights (Harris-Warrick & Marder, 1991). Most neural network simulations in cognitive science gloss over these fine-grain details and use relatively simplified activation functions and connectivity patterns. Including all of those biologically precise parameters into a simulated neural network and having billions of nodes with trillions of connections (like the human brain has) will require truly massive computing capacity (McClelland, 2009).

Second, most artificial neural networks exhibit a form of **convergent dynamics**. The activity of such a network eventually dies down and reaches a stable state. This is not the case for brain activity. Real neural networks are oscillatory and chaotic—their states fluctuate over time and do not settle. Neural activity in the hypothalamus, for example, is oscillatory. Regarding this type of dynamic, it is a fascinating but difficult question to determine where in the network a representation exists. In convergent networks, the presentation of a stimulus pattern reliably elicits the trained response. This response is the network's distributed representation of the correct answer in the form of a pattern of activation among nodes. But in oscillatory and chaotic networks, there is no such stable representation. The network, because it is constantly changing, has not formed any specific pattern of activation that can be linked to a stimulus. Representations in these systems might correspond to more global characteristics, such as the frequency or phase of the network's activity (Skarda & Freeman, 1987).

A third problem that arises in network learning is the **stability–plasticity dilemma**. It states that a network should be plastic enough to store novel input patterns; at the same time, it should be stable enough to prevent previously encoded patterns from being erased (Grossberg, 1987). This apparent conflict is analogous to the phenomenon of psychological interference, discussed previously. The fact that ANNs show evidence of being caught in this dilemma is useful because it may offer some insights into human interference. But it becomes a real problem when one is attempting to implement artificial networks. **Catastrophic interference** occurs in instances in which a network has learned to recognize a set of patterns and then is called on to learn a new set (French, 2001). The learning of the new set modifies the weights of the network in such a way that the original set is forgotten. In other words, the newly learned patterns suddenly and completely ("catastrophically") erase the network's memory of the previous patterns. A variety of solutions to the problem have been proposed, ranging from the introduction of "novelty vectors" to a reduction in the internal overlap of representations, to pretraining the network with structured samples of patterns (French, 1992; Kortge, 1990; McRae & Hetherington, 1993).

Finally, in backpropagation networks, the error signal is propagated backward through *multiple* previous layers of nodes. However, there is no evidence that biological networks can feed an error signal (for modifying connection strengths) any further back than the immediately preceding layer of neurons (Bechtel & Abrahamsen, 1991). For the backpropagation learning algorithm, this is a convenient and powerful engineering solution to the problem but fails to address how an analogous process in actual brains might occur.

SEMANTIC NETWORKS: MEANING IN THE WEB

There is another class of network-like models that have some features in common with neural nets. Nodes, links, thresholds, and summation of inputs also characterize these networks. They are **semantic networks**. But in semantic networks, there is no distributed coding. Each node has a specific meaning. Semantic networks, therefore, employ the localist representation of concepts. Semantic networks were adopted primarily by cognitive psychologists as a metaphor for describing the organization and retrieval of information in long-term memory. In this section, we will discuss the characteristics of semantic networks, drawing on experimental results from cognitive psychology.

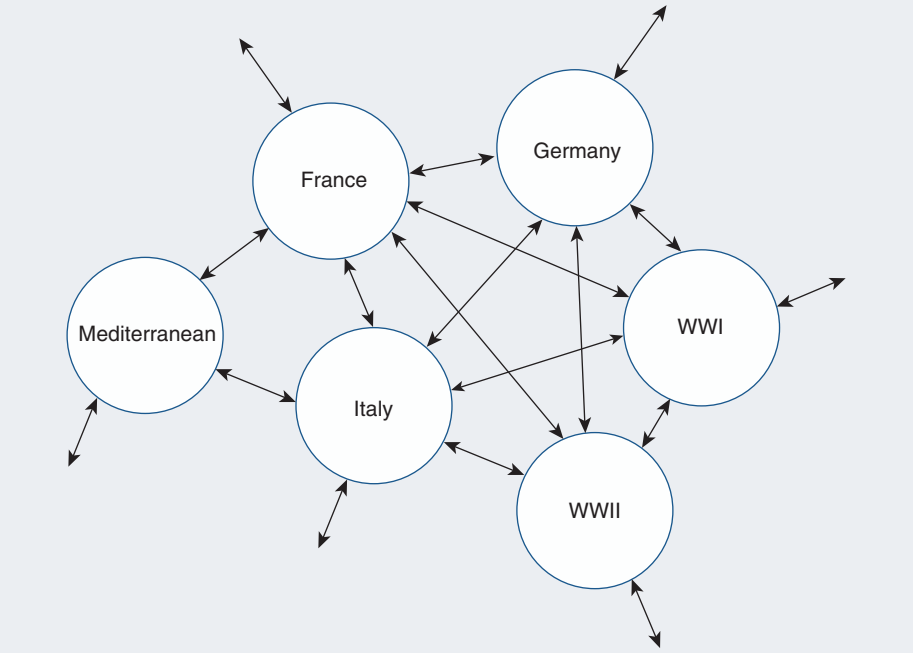
One motivation for using semantic networks centers on representational transparency. The distributed representations formed in ANNs are often difficult to analyze and identify. The distributed coding in neural networks results in each node participating to some degree in multiple different representations. With semantic networks, however, each node is labeled in advance to represent a specific unique concept or feature. Moreover, when there are different types of links between nodes, it can allow semantic networks to represent a variety of types of relationships between concepts and features.

Characteristics of Semantic Networks

In semantic network models, a node's activity can spread outward along links to activate other nodes. These nodes can then activate still others—a process called **spreading activation**. An assumption of some models is that activation energy decreases with increasing distance, since it is believed that the activation energy encounters resistance as it passes through succeeding links and nodes. This means that spreading activation may lose strength as it travels outward from its point of origin. Another characteristic of these networks is that the number of links to travel from one node to another is determined by their degree of relatedness. Concepts such as "automobile" and "truck" are semantically related and, hence, will have a direct connection between them. By contrast, concepts such as "automobile" and "skateboard" are less related and thus would only be indirectly connected via a few intervening concepts (e.g., with motorcycle and bicycle in between them). Therefore, if the "automobile" node were activated, it would be faster and easier for activation to spread to the "truck" node than to spread to the "skateboard" node.

Spreading activation is thought to underlie retrieval of information from long-term memory. For example, suppose a test question in history class asks for the names of major western European nations. While you were studying, you had memorized these nations in several ways: on the basis of their geographic locations, whether they fought each other in wars, how related their languages are, and so on. You start by naming Germany, then Italy. At the end, you realize you've left out a country but can't think which it is. Finally, it comes to you. You remember that the country is France—because while you were studying, you realized that France and Germany fought each other during the two world wars. Although you were not able to retrieve France on the basis of the Mediterranean characteristics it shares with Italy, you were able to retrieve it on the basis of its historical relation with Germany. This underscores an important principle of effective studying: It is best to associate a concept with as many other related concepts

Figure 7.9 A hypothetical example of how knowledge of western European countries might be represented in a semantic network.



Source: Adapted from Sejnowski and Rosenberg (1986).

as possible, to ensure a higher likelihood of retrieval. In this way, there exist a greater number of alternative pathways that lead to a given node. If one nodal connection fails to activate the target item, spreading activation will ensure that some other pathway does. These alternate associations that facilitate recall are also called **retrieval cues**. Figure 7.9 illustrates how a semantic network might represent knowledge of western European countries.

An important characteristic of semantic networks is **priming**. In priming, the processing of a stimulus is facilitated by the network's prior exposure to a related stimulus. A study by Meyer and Schvaneveldt (1971) demonstrates this phenomenon. The participants in their study were presented with pairs of letter strings. Each letter string was either an actual word or a nonsense word. Participants were asked to push a button when they encountered actual words. The button push recorded how long it took them to respond. If both words in a pair were meaningful and related, the participants were faster to respond correctly to the second word in the pair. So if the first word was *nurse* and the second *doctor*, responses were faster than if the first word was *nurse* and the second *butter*.

The explanation goes as follows. The "nurse" node in long-term memory becomes activated first. Through spreading activation, it activates other, semantically related, nodes, such as "doctor." A short time later, when the participant's response to *doctor* is required, the response time is reduced because that node is already partially activated. This technique can be used to map out the structures of semantic networks. The larger the priming effect, the more closely connected the relevant nodes should be (McNamara, 1992).

A Hierarchical Semantic Network

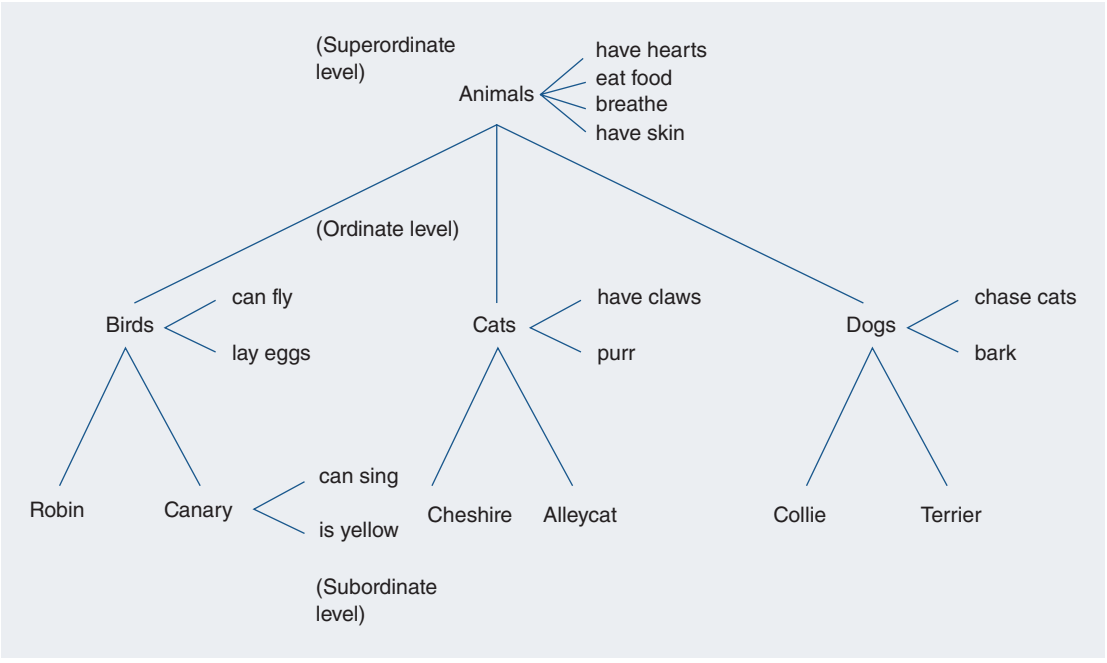
A classic study by Collins and Quillian in 1969 suggests that semantic networks may have a **hierarchical organization**, with different levels representing concepts ranging from the most abstract down to the most concrete. They used a **sentence verification** task. Participants were asked to respond to individual sentences that appeared on a computer screen. If they judged a sentence to be true, they hit one button. If they judged it to be false, they hit another. Examples of true sentences were "A canary can sing" and "A canary has skin." Response times were recorded.

Collins and Quillian (1969) theorized that a correct response to one of these sentences required overlap of spreading activation. Verifying "A canary can sing" would require the activation of both the "canary" concept node and the "can sing" property node. Activation would then radiate outward through the network from each node until individual units' activations would mutually affect one another. When the activation of these two nodes overlaps, the participant knows that the two are related and can then verify the sentence as true. If the nodes have a close semantic relation, they should be in proximity to each other in the network, and responses will be fast because spreading activation will have less distance to cover. If the nodes are less closely related, such as "canary" and "has skin," the distance between them will be greater and response times will be longer.

Based on the set of reaction times they obtained in the study, Collins and Quillian (1969) sketched out a hypothetical memory structure for knowledge of animals. This is shown in Figure 7.10. The hierarchical nature of the network is immediately clear. The concept "animals" and its properties—such as "eat food," "breathe," and "have skin"—are at the top of the hierarchy. Because the idea of animals is abstract and encompasses all known types of animals, *animals* constitute a **superordinate** category. At the next lower level, we have examples of classes of animals—such as "birds," "cats," and "dogs"—accompanied by nodes corresponding to their characteristics. For birds, these would be "can fly" and "lay eggs." Because these different classes of animal are at an intermediate level of specificity, they are **ordinate** categories. At the bottom of the hierarchy are nodes that are even more concrete, corresponding to exact species of animals; these animal species are **subordinate** categories. Properties of these animal species are also at this same level. A node corresponding to "canary" would have links at the subordinate level connecting it to properties that are relatively specific to it, such as "can sing" and "is yellow."

Figure 7.11 depicts a graph of the reaction times from Collins and Quillian (1969) that led to the model being proposed. Responses for which spreading activation occurs

Figure 7.10 The hierarchical semantic network proposed by Collins and Quillian (1969).

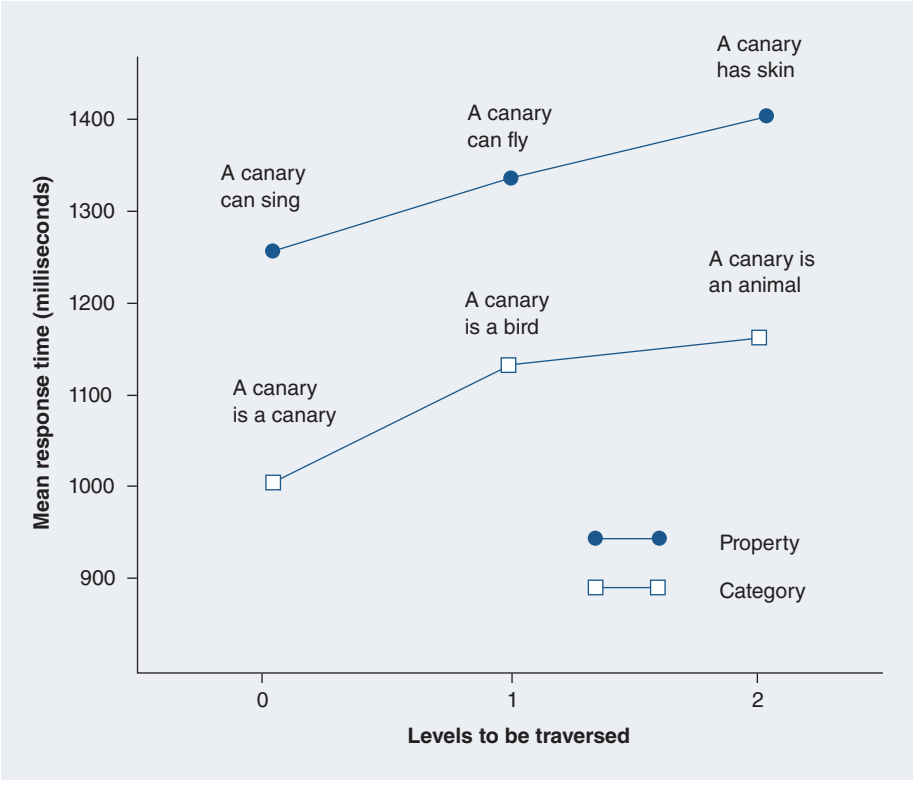


Source: Adapted from Collins & Quillan (1969).

entirely within a level were quite fast. It took participants about 1.25 seconds, on average, to confirm that "A canary can sing." In this instance, spreading activation did not have far to travel, connecting the "canary" node to its attached property node, "can sing." By contrast, to confirm "A canary has skin" took slightly longer—about 1.4 seconds. Presumably, the longer time is owing to the fact that the "canary" node (at the subordinate category level) is rather distant from the "has skin" property node attached to the "animal" concept node (at the superordinate category level).

Finally, you will notice that response times for category statements follow this same function. Sentences like "A canary is a canary" were quickly verified, whereas sentences like "A canary is a bird" took slightly longer because activation has to spread from the subordinate category level to the ordinate category level. When activation has to spread from the subordinate category level all the way to the superordinate category level, as in "A canary is an animal," reacting times were longer still. In general, category statements were faster than property statements because activation need not spread from a concept node to a property node when simply verifying categories.

Figure 7.11 A graph depicting the results of the Collins and Quillian (1969) experiment.



Source: Adapted from Collins & Quillan (1969).

Evaluating the Hierarchical Semantic Network Model

The hierarchical semantic network model is intuitively appealing but fails to account for several findings (McCloskey & Glucksberg, 1978). For example, participants are faster to make the correct response to "A dog is an animal" than to "A dog is a mammal," even though "animal" is farther away from "dog" in a hierarchical scheme. Evidently, people tend to think of dogs more as animals than as mammals, even though "mammals" is a subcategory of "animals" and would be closer to the concept of dog than "animals" is. Also, the sentence "A robin is a bird" is responded to more quickly than "An ostrich is a bird," even though the two sentences are equivalent and ostriches and robins are both subordinate to the bird category. This suggests that some birds are more "typical" than others and that concepts may be represented by **prototypes** that represent generic or idealized

versions of those concepts (recall the discussion of category typicality from Chapter 5). If this were the case, then the network would need restructuring, and the "robin" node would have to be moved so that it was closer to "bird" than was the "ostrich" node. These results imply that the hierarchical organization of the sort seen in the Collins and Quillian (1969) model may be oversimplified and may represent only one of the ways in which we organize concepts.

A further criticism of the hierarchical model has to do with the principle of **cognitive economy**. This principle states that nodes should not have to be coded for more times than is necessary. Collins and Quillian (1969) attempted to preserve cognitive economy by placing property nodes only at their appropriate levels in the hierarchy. For instance, "can fly" is positioned as a property of birds only, even though it could be linked to the canary node and to nodes representing every other kind of bird as well. The advantage of cognitive economy is obvious. It eliminates redundancy and frees up resources in a memory system. But research has found no difference in the response times for the sentences "A bird has feathers" and "A canary has feathers." So the property "has feathers" seems to be attached to our concept of "canary" as well as to that of "birds." Like the network hierarchy, cognitive economy seems to work better in principle than in reality.

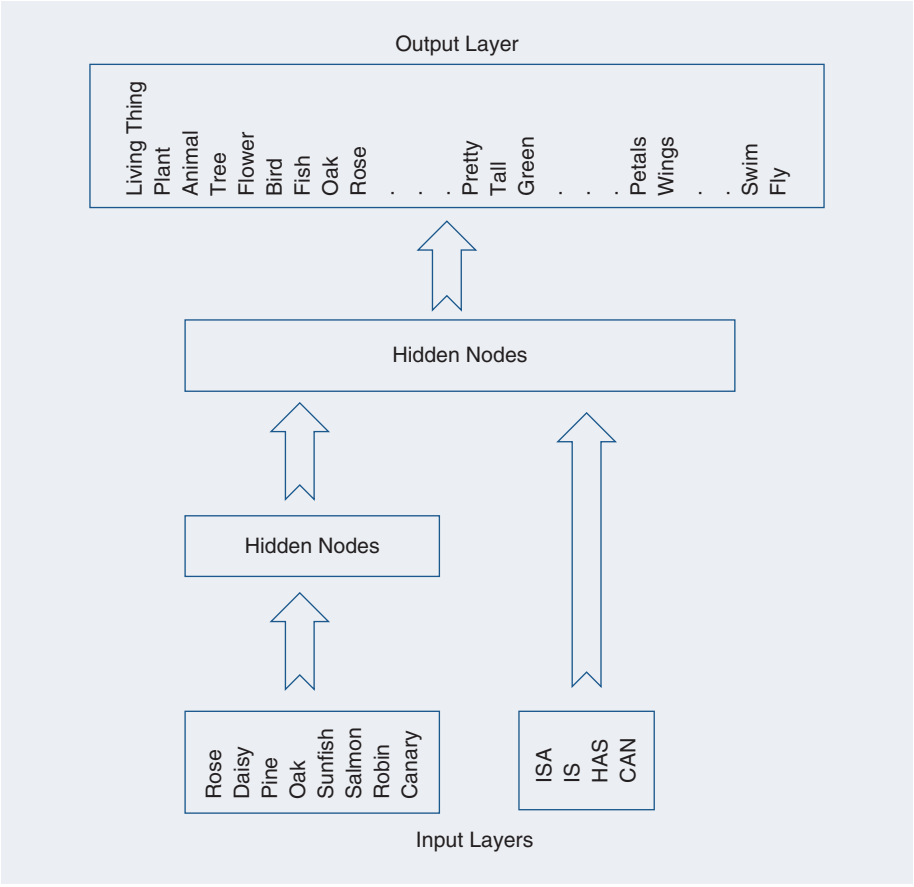
Finally, an important weakness of the hierarchical semantic network is that it is clearly a metaphorical description of some form of knowledge representation that is actually encoded by neurons in the human brain. A hierarchical structure like that in Figure 7.10 is not literally inside the human mind. Therefore, the semantic network fails to tell us how category and subcategory knowledge is encoded in a brain. To explore this, McClelland and Rogers (2003) extended a backpropagation model first developed by Rumelhart and Todd (1993). The hidden layers of this neural network (see Figure 7.12) developed clusters of data for the categories with subclusters of data for the subcategories, simply by being trained on thousands of sample input–output pairings (e.g., Robin+ISA = Animal, Rose+HAS = petals). Without using any explicit hierarchy in its knowledge structures, the network nonetheless learned these relationships in a way that was equivalent to hierarchical knowledge. Moreover, McClelland and Rogers were able to document the gradual development of these knowledge structures at intermediate stages of learning in the network and found that its early errors are in fact similar to the kinds of errors that young children make with category knowledge.

Evaluating Semantic Networks

Semantic network models can be implemented as computer simulations and then put to the test to see if they account for humanlike memory processes. This has been done, and although many of the simulations can reproduce aspects of performance in limited domains, there is no semantic network model that successfully simulates the full variety of human knowledge.

A phenomenon that presents a challenge to semantic network models is called the **T.O.T. phenomenon**, the acronym standing for "tip of the tongue." Try to recall the name of the navigational device used by sailors a century ago that measures the positions

Figure 7.12 **A schematic diagram of the Rumelhart and Todd (1993) model that McClelland and Rogers (2003) extended to demonstrate the development of category knowledge.**



of the stars to determine longitude and latitude. You may find yourself frustrated in trying to recall the word, as many people can come up with the first letter and the number of syllables but not the word itself (Brown, 1991). A person who gets stuck trying to remember the word is, in a manner of speaking, in the vicinity of the node or network that codes for the word because he or she has some knowledge of the word. This person also presumably has activated nodes that are in proximity to that of the concept in question because he or she is devoting a great deal of time and effort to recalling the word. So why doesn't spreading activation activate the appropriate node? Semantic

networks cannot easily explain these sorts of retrieval blocks and the availability of partial information such as the first letter and the number of syllables. By the way, if you couldn't get it, the word is *sextant*.

The phenomenon opposite to the T.O.T. phenomenon is the situation in which we can successfully retrieve an item from memory despite the fact that there are no close connections between retrieval cues and the target. Reisberg (2001) gives a good example when he asks us to think of the node that represents "water." Because water is an eminently familiar concept, it has connections to many others. The water node thus ought to have multiple links that radiate outward toward other nodes—a high **degree of fan**. In contrast, the concept "xylophone" has a low degree of fan, at least for most people, as we don't often think about this musical instrument. Assuming that the water node has links to 100 other nodes—probably a conservative estimate—then spreading activation would, in just one step, activate all 100. Assuming that each of these 100 nodes is, in turn, connected to 100 other nodes, then 10,000 (100 × 100) nodes would be activated in a memory search after only two steps. The number of activated nodes increases tremendously with increasing numbers of steps. This process is inefficient from a computational standpoint. It is also counterintuitive. It certainly doesn't feel as though we think of a million or more things when we are trying to remember the capital of Texas.

A solution to the problem of excessive activation would be the implementation of an inhibitory network. Here, the links are inhibitory, so activation of a given node causes it to dampen the activity of its neighbors. In this model, the greater a node's activity, the more it will inhibit the nodes to which it is immediately linked. Nodes that receive several inputs are more highly activated and thus stand out in the crowd since they can suppress their neighbors more powerfully. Concepts with multiple inputs are probably the ones we are searching for because these input pathways are what are activated during the retrieval process. To go back to our earlier example, if we are trying to remember "France," this node ought to stand out among the others in an inhibitory network. This is because it is activated in two ways: (1) on the basis of the "commonly taught languages" path and (2) on the basis of the "fought with Germany" path. Inhibitory networks thus dampen activity in nodes that are unlikely candidates in a memory search and increase activity in nodes that are likely candidates.

One of the nice things about semantic networks is that the search for an item takes place automatically. All one needs to do is hear a question, and the answer seems to pop into one's head. The relative ease with which we recall, at least much of the time, is in accord with the idea of spreading activation. The question triggers nodes that are near the desired node; activation then spreads across the network until the target item is sufficiently activated. But, as we all know, recall is not always this easy. Sometimes, as in the T.O.T. phenomenon, we must expend effort trying to remember something. If somebody asked you what you did on your last birthday and it didn't immediately come to mind, you might engage in a **guided search**—one governed by intelligence and reasoning. You would think of your friends and members of your family who might have been at the party, what kinds of presents you received, and where the party was held. All these could serve as retrieval cues or associated nodes and, as such, are elements in the network. But they weren't triggered automatically. Instead, you had to willfully and

deliberately reconstruct what happened on your last birthday, based on what you know of your past birthdays in general. This type of **reconstructive memory** is used quite often and has been shown to be influenced by environmental cues, such as how the question is phrased (Loftus, 1979). But the important point here is that reconstructive memory constitutes a separate process of retrieving items—one that does not rely solely on spreading activation and the inherent, automatic characteristics of the network.

NETWORK SCIENCE

The past couple of decades have seen the development of a new area of study called **network science** (Watts, 2003). Scientists, mathematicians, and others in this area have begun to explore the way in which complex networks of all kinds operate. A network in this view is considered as any collection of interconnected and interacting parts. Like cognitive science, which is made up of investigators from multiple disciplines, network science is also interdisciplinary. Investigators come from different areas such as mathematics, physics, finance, management, and sociology. This is because networks are everywhere, in the form of electromagnetic fields, stock markets, companies, societies, the Internet, and, of course, the brain.

Contemporary network scientists additionally consider networks as dynamical systems that are *doing* things. Power grids distribute power, companies manufacture products, and brains think. All these are networks in action. The structure of networks can also change over time. New links can form and new organizations emerge. The new science of networks investigates not just network structure but also its function. Researchers in this field are interested in the anatomy of networks, how they operate, and how these operations evolve and change.

Network scientists consider networks as abstract structures, and to some extent, they can disregard their specific context. Whether the network is a brain, a traffic system, or the Internet, in some sense, doesn't matter. That is, because there are commonalities these networks share. In other words, all networks share some universal mechanisms of action. For instance, most known networks exhibit a critical point where activity propagates suddenly through the entire system. It seems amazing that cars on the road and employees in a company should act in the same way, but, because they are both elements in interconnected relationships, they can.

It is also important to note that although networks do share some characteristics, they don't act exactly the same. In some cases, different types of connectivity are responsible for producing different behaviors out of the system. For example, some networks are minimally connected, and most of the work is being performed by operations inside the nodes (or components) themselves, exhibiting **component-dominant dynamics** (Van Orden, Holden, & Turvey, 2003). By contrast, other networks are densely interconnected, and most of the work is being performed by the connections themselves (not inside the nodes), thus exhibiting **interaction-dominant dynamics**. Networks with interaction-dominant dynamics (due to their dense interconnectivity) often produce unexpected behaviors that are difficult to trace back to their causes. These are called **emergent properties**. In

other cases, the nature of the nodes that make up the network is what produces different behavior. A neuron in the brain acts differently from a consumer in a society. So we need to acknowledge that there are both universal and particular features to networks. In the following sections, we describe aspects of network science and how they apply to cognition. Friedenberg (2009) provides a more detailed treatment of these topics.

Centrality

One of the most important questions concerning networks is how activity in them is coordinated. A network might have one or perhaps several centers that perform this function. This is known as the issue of **centrality**. If there is a "leader" or part of the network that receives information, evaluates it, and issues commands, then centrality is accounted for. Many examples of such systems exist. Computers have a central processing unit, armies have generals, and political organizations can be run by a dictator. The more interesting case is how networks without any such center achieve this kind of coordinated action.

This question has particular relevance for the human mind. For many years, it was believed that the brain must have a central coordinating agency, a center where everything comes together. This was the notion of the Cartesian theater that we addressed in the philosophy chapter. This center begs the question, though, because if we allow for a homunculus, then we must explain how it coordinates. If we cannot, then we are stuck with the familiar infinite regress of postulating a homunculus inside the homunculus's mind, and so on.

When we introspect, it seems that we have a single unified self. This seems to run contrary to the known physical evidence that activity is occurring throughout many separated brain areas simultaneously. Some have suggested that consciousness must be the mechanism by which this distributed activity is unified into a single form of awareness. If we could figure out the centrality issue, we might also determine the answer to the mystery of consciousness. Candidate explanations for this include Dennett's (1991) multiple drafts model and the phenomenon of neural synchronization (Singer, 1996).

In some networks, coordinated global activity happens simply as a function of spreading activation that disperses throughout the system quickly but can arise from any part of it. Take, for example, the clapping of an audience after a show. If one group of people starts clapping in synchrony, the sound gets louder than the clapping in their immediate surround. This causes the audience nearby to clap in synchrony with them, which raises the group's volume even more and perpetuates its spread until the entire audience is clapping in lockstep at the same speed (Néda, Ravasz, Brechet, Vicsek, & Barabási, 2000). What makes this interesting is that it can originate in any part of the audience. All that is required is a minimum number of initial starting clappers as a seed to start the propagation.

Hierarchical Networks and the Brain

Connections in **hierarchical networks** are organized into different levels. The number of nodes decreases as we travel upward, with each group of nodes converging

onto another group above it with fewer elements. We already have seen hierarchical networks. Figures 7.4 and 7.5 are examples. Hierarchical organization is perhaps most evident in the visual system (Hubel & Wiesel, 1962). Here, neurons at the bottom of the hierarchy code for specific features, such as lines at a particular orientation and at a specific position in the visual field. Connections from these **simple cells** converge on neurons in the next layer up. These cells code for the combined input of the lower features that feed to them. If they received inputs from multiple adjacent vertical line detectors, they would in turn code for the sum of these representations, which would be a vertical line moving in a specific direction. These **complex cells** next converge to neurons at an even higher level that would again stand for their combined inputs. In this case, these new **hypercomplex cells** represent a right angle made of vertical and horizontal lines that are moving conjointly in the same direction (Figure 7.13).

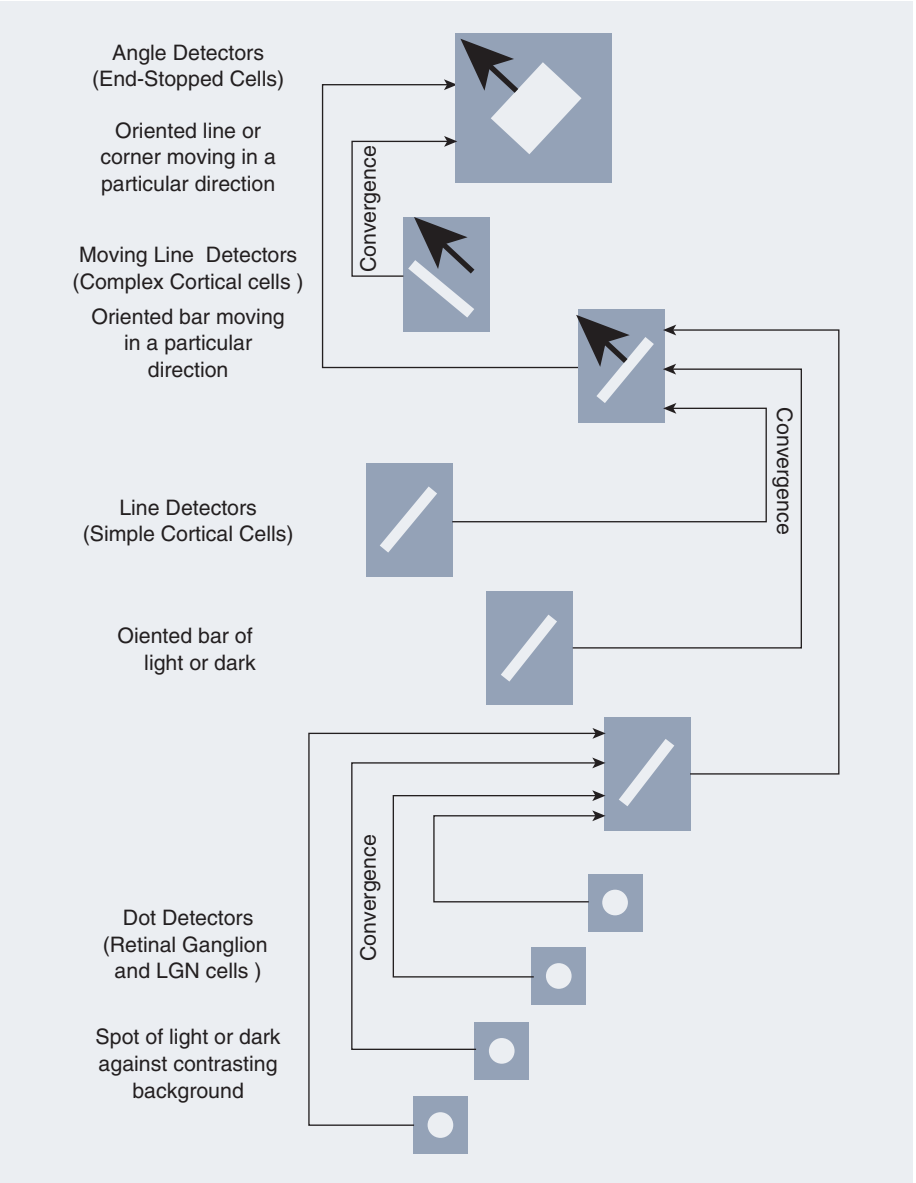
If we extrapolate up the hierarchy, we end up with small ensembles of cells in the highest layers that code for large complex objects. (Recall the discussion of sparse distributed coding in Chapter 6.) So the visual system uses hierarchical networks to successively integrate features of the visual field and build a percept. In effect, the hierarchy allows the visual system to employ a "divide-and-conquer" strategy where it breaks down the complex visual image into microscopic features and then assembles these features into parts and then wholes that can be recognized. Similar hierarchies may exist that allow concepts to be formed from component characteristics. This example illustrates an important feature of hierarchies. Traveling up through them allows for convergence and integration—that is, the putting of pieces together. Traveling downward through them allows for divergence and separation (a splitting apart or decomposing of elements) and perhaps also for making predictions of future input (Rao & Ballard, 1999).

Communication between levels in hierarchies can also allow for the resolution of ambiguity in visual perception. Recall the word *superiority effect* discussed in the first cognitive chapter. In this effect, it is easier to determine the identity of a feature—say, the letter *I*—when it is presented as part of a meaningful word, such as *fit*, than when it is presented by itself as an isolated letter. The surrounding letters in the word provide a context that helps disambiguate the central letter's identity. But this could happen only if information about the word, processed at a higher level, was compared against information about the letter, processed at a lower level. This shows that perception is not purely "bottom-up" or data driven, with only parts forming wholes. It must also be "top-down" and conceptually driven, with wholes helping form the developing percept. Information in the visual system thus appears to travel in two directions. It goes not only in a feed-forward direction from the eye to the brain but also in a feedback direction from higher brain centers to lower centers. Both are necessary parts of the pattern assembly process.

Small-World Networks: It's a Small World After All

A number of years ago, several undergraduate students at Albright College discovered an interesting fact. They realized that every American movie actor could be related to the actor Kevin Bacon in only four steps or less. The links are through actors who appeared in the same film together. The actor Will Smith, for example, was in the movie

Figure 7.13 Hierarchical organization in the visual system. Converging input from lower layers results in the coding of more complex objects.



Source: Friedenberg (2009).

Independence Day (1996) with Harry Connick Jr., who was in the film *My Dog Skip* (2000) with Kevin Bacon. So in this case, only two steps separate Kevin Bacon from Will Smith. This finding garnered widespread public attention on television and the Internet. One website, called the Oracle of Kevin Bacon, received 20,000 hits a day at its peak and was ranked as one of the top 10 websites of 1996.

This finding is real but it is not specific to Kevin Bacon. Any two well-known American actors can be connected by a maximum of four links. In fact, the principle holds not just for film actors but also for most social relationship networks. The psychologist Stanley Milgram first noted this when he sent out 160 letters to people. The letters were addressed to a stockbroker. He asked these people to pass on the letter to someone who they thought would know the stockbroker. Surprisingly, a number of the letters did actually get through, most of them in about six steps (Milgram, 1967). This small-world network is actually far more general than even social relationships. It exists in the U.S. electrical power grid, in road and railway networks that are spread out across continents, and even in the nervous systems of many animals (Van den Heuvel, Bullmore, & Sporns, 2016).

So we can define a **small-world network** as any network where one can get from any single point to any other point in only a small number of steps even though the total number of elements may be exceedingly large. How could the units of such networks be so easily connected when the networks themselves are so vast? After all, there are billions of webpages and billions of people on Earth. It seems as though it would take a large number of connections to link any two elements together. There must be some intrinsic organization or structure to these networks that allows such ease of connectivity. We examine this issue next.

Ordered and Random Connections

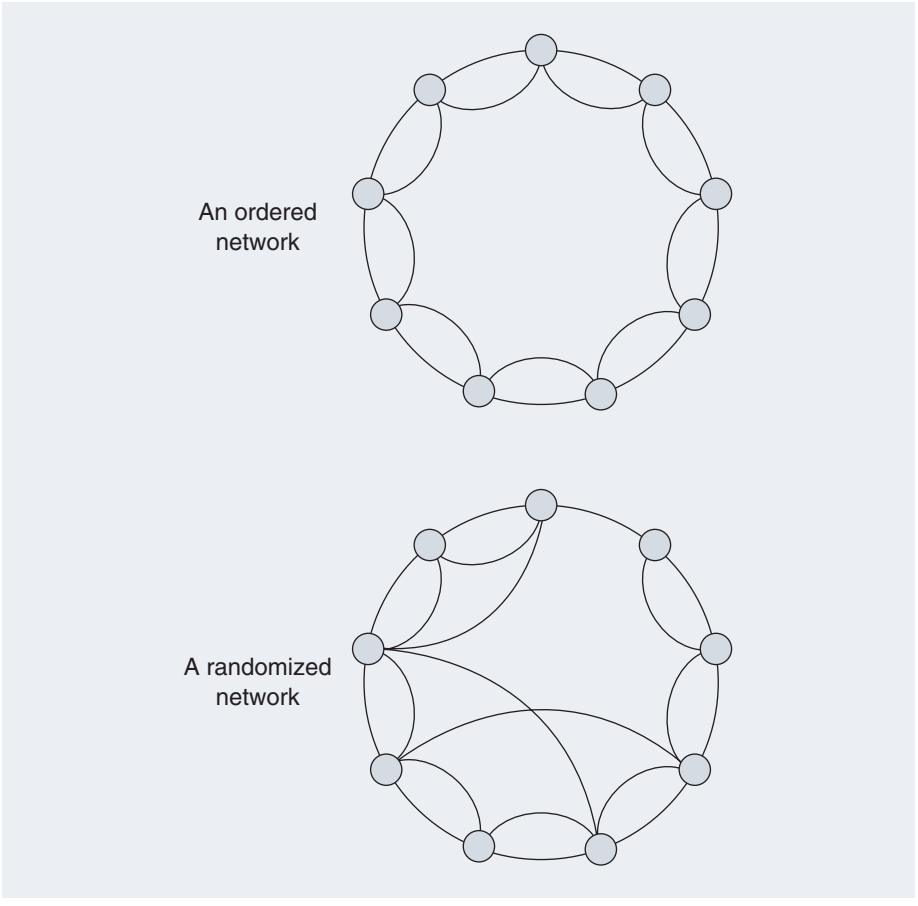
The answer to this question has to do with network architecture. Watts and Strogatz (1998) simulated a variety of network types. At one end, they created a highly ordered network with only local connections. In this network, a node was connected only to other nodes in its immediate neighborhood. At the other end of the continuum, they had a completely random network where a node could be linked with any other node in the network with equal likelihood. **Random networks** of this sort contain an equal mix of link distances, both local and global.

Let's start with a network of 6 billion nodes and imagine them arranged in a circle (Strogatz, 2003). The population of the world was 6 billion several years ago, so we can imagine each node standing for an individual person and the links between nodes standing for a type of relationship. In the case of an **ordered network**, each node might be connected with 1,000 others that are close by—500 to the left and 500 to the right. Each connection can be represented as a line drawn between nodes that are themselves denoted as smaller circles (Figure 7.14). In this ordered network, it would take 3 million steps on average to link any one node to any other. The farthest connection is a node diametrically opposite its counterpart on the circle. To reach it would require hopping around the circle in steps of 500. It would thus take 3 billion divided by 500, or 6 million, steps to reach. Since the closest node is just one step away, the average number of steps

separating any two nodes would be midway between 1 million and 6 million, a whopping 3 million.

Now, let's take the case of a completely random network. Here, it takes only four steps to get between any pair of nodes, on average. If every person knows 1,000 others (randomly located on the large circle) and each of those knows 1,000 additional people, then there are 1 million people ($1{,}000^2$) within two steps of each other, 1 billion ($1{,}000^3$) within three, and 1 trillion ($1{,}000^4$) within four. So any two people in this system are likely to be separated only by three or four degrees—just like Kevin Bacon.

Figure 7.14 Ordered networks have only local connections to immediate neighbors. Random networks have more long-distance connections that allow messages to travel faster through the network.



Source: Friedenberg (2009).

Egalitarians and Aristocrats

It turns out that there are at least two kinds of small-world networks. The first is of the type discovered by Watts and Strogatz (1998), where a few random connections are thrown in among mostly local neighborhood links. Buchanan (2002) calls these **egalitarian networks** because the links are distributed more or less equally. But there is also what he calls **aristocratic networks** (or scale-free networks; Barabási, 2009). Here, the long-distance connections result from the presence of a small number of hubs. **Hubs** are nodes that have many more links than the rest of the network's nodes. There are numerous examples of hub-based aristocratic networks, both in the natural and technological realm. Examples include the Internet and the World Wide Web, the spread of sexual disease among people, scientific papers linked by citations, and the probability of words occurring next to one another in English sentences.

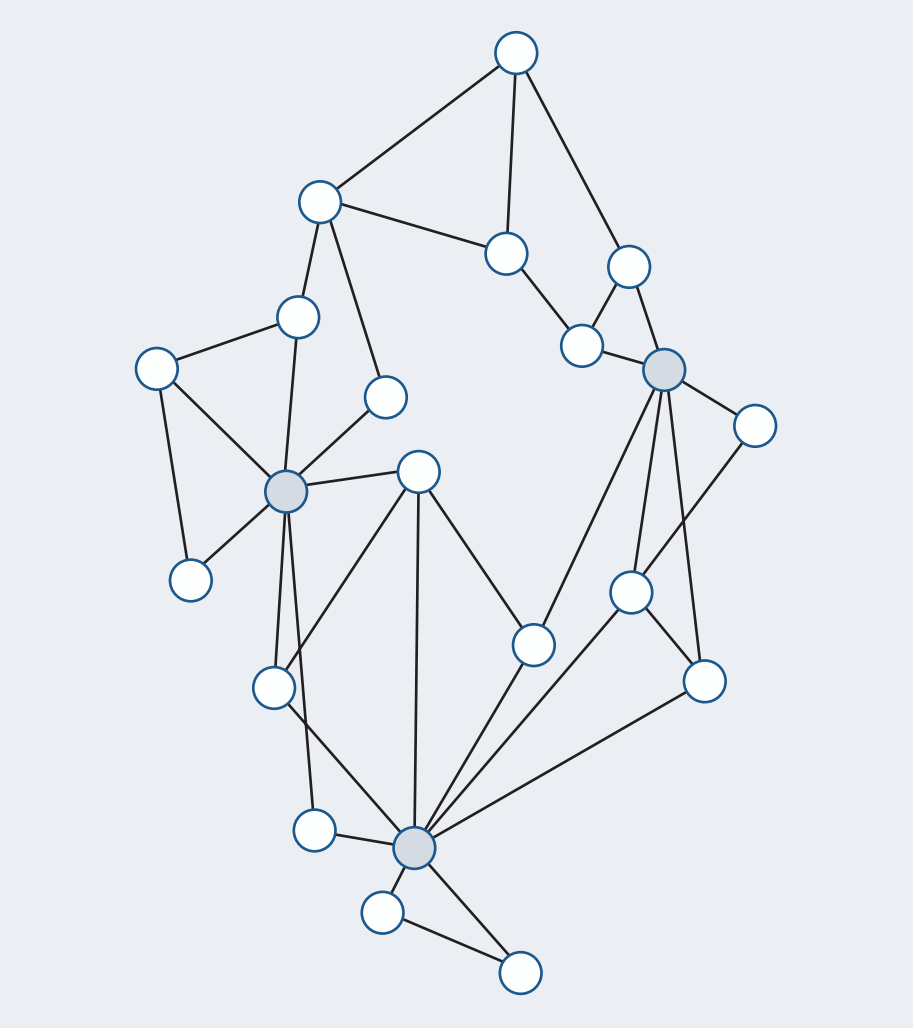
The aristocratic organization seems to emerge through a very simple process, one that Buchanan (2002) calls "the rich getting richer" or, more properly, **preferential attachment**. When a network is created, there can be slight differences in the number of links each node has. Those with more links will tend to be the ones that acquire brand-new links more often because they serve as good portals to gain access to large numbers of other nodes. If you created a website on the World Wide Web and wanted lots of people to access it, you would link to other popular sites such as Yahoo.com or Google.com, because you know that lots of other people are already visiting those pages. Hubs, therefore, turn the network into a small world, one where information flow between widespread points is facilitated. The fact that networks of all kinds naturally and unintentionally develop hubs as they grow suggests that it is an emergent property, a solution nature has found to solving the problem of rapid and efficient communication.

The two types of small-world network both accomplish the same function but in different ways. As a result, they each have different weaknesses (Baran, 1964). Egalitarian nets are more robust to nonspecific damage. They can stand up better to the destruction of randomly selected nodes. As long as there are several long-distance connections intact, messages can get rerouted along them. Aristocratic nets are more sensitive to selective attack where the hubs are targeted and destroyed. Because the hubs carry most of the links, the destruction of only a few can shut down the entire network. The current Internet (of connected computers) and World Wide Web (of connected websites) are both aristocratic, or scale-free, networks. This means that special attention should be given to securing those hubs and shielding them from attack. Figure 7.15 shows an example of a hub-based network.

Neuroscience and Networks

Let's now apply these ideas to cognitive science. One question we can ask is whether the brain is a small-world network. Scannell (1997) studied the cortex of cats and monkeys. He located 55 distinct cortical regions in the cat, each associated with a particular function. In macaque monkeys, there are 69 such regions. For both of these animals, about 400 to 500 links connect these areas. These links consist of not only single axons but also nerve fibers made up of numerous axons. Latora and Marchiori (2001) analyzed

Figure 7.15 An example of a hub-based network where some nodes have many more connections than others. The three hubs are indicated by shading.



Source: Friedenberg (2009).

Scannell's data and determined that there are only between two and three steps separating neurons. They also found that these neurons were highly clustered, with many shared local connections.

So it appears that mammalian brains are small-world networks after all. This makes sense. If an animal is to respond rapidly to a stimulus—say, to the sight of a predator or

prey—the information can't pass through too many synapses. Each synaptic link would add additional response time, slowing down the animal and putting it at risk.

Epilepsy is a brain disorder characterized by uncontrolled neuronal firing. Much like a forest fire that can start in one part of the forest and then spread, epileptic seizures can also multiply. Unfortunately, the brain's long-range connections allow a seizure in one part of the brain to easily spread to other areas. It is not uncommon for a seizure to expand from one hemisphere to another along the corpus callosum or from one cortical area in one hemisphere to another that is more distant but on the same side.

In an ischemic stroke, an object such as a blood clot blocks blood flow and kills neurons by denying them oxygen and nutrients. Depending on the location and extent of damage, several kinds of disorders can result. Many stroke patients, however, can recover all or most of their former ability through therapy. What is happening is that their brains are reconfiguring, in some cases forming new synaptic pathways that avoid or flow around the area of damage. The ability of the brain to recover from injury in this fashion is called **plasticity**. It is the dense weave of local connections that allows this reconfiguration to occur. If one node or set of nodes is knocked out by the stroke, the signal passing through there can reroute along alternate intact pathways. **Clustering** thus ensures that the brain doesn't fragment into isolated islands.

Small-World Networks and Synchrony

In the previous chapter, we discussed the importance of neural synchrony, which generally refers to correlated activity between neurons. It turns out that small-world networks seem to promote **synchrony**. Kuramoto (1984) found that synchrony between coupled oscillators with distributed natural frequencies (each firing at a variety of different rates) failed to occur in ordered networks, where each oscillator was connected only to its immediate neighbors. But randomizing this network just a little bit produced global synchrony, where all the oscillators beat in unison (Strogatz, 2003). The explanation is intuitive: Just a few long-distance connections transmit timing information to more distant parts, allowing mutual influence to spread quickly. The fact that small-world networks and synchrony are both properties of brains is no accident. The former is a prerequisite for the latter.

Lago-Fernandez, Huerta, Corbacho, and Siguenza (2000) created a simulation of the olfactory antennal lobe in the locust. They connected the neurons in their model together in different ways and measured how the network responded to simulated olfactory input representing smells. When the network was ordered with only local connectivity, the simulated neurons synchronized but took a long time to do it. This sluggish response was ostensibly due to the time it took for the timing signals to propagate throughout the system. Next, they wired up the network at random. Now, the signal spread rapidly throughout the network, but synchrony didn't occur. The randomly wired networks apparently lacked sufficient order to allow the network to produce an organized response.

Finally, these researchers tried ordered networks with just a slight amount of randomization. Now, the system synchronized and did so quickly. But Lago-Fernandez et al. (2000) discovered another interesting finding in addition to this. In real-world locusts, some of the neurons either slow down or speed up compared with the group frequency.

They speculate that this frequency distribution codes for odor intensity or for different types of odors. Only the small-world wiring pattern was able to replicate this effect.

Percolation

If we wish to better understand how information spreads in a network, we can look at the spread of disease. A disease starts off by infecting one or more individuals, who then come into contact with others. These new people become infected and infect yet additional people. A number of factors affect how far a disease will spread in a social network. The interconnectedness of the network is important, as is the susceptibility to infection. An elementary school teacher can contribute greatly to the spread of a disease because he or she comes into contact with many children, who may all be highly susceptible to infection. A loner who works at home most of the time, on the other hand, may not infect anyone at all.

A number of models have been proposed for how disease spreads. Here, we will discuss the percolation model. The term **percolation** refers to the propagation of a disease through a social network. We start with a large population of people connected by bonds along which the disease can spread. Each site is either susceptible or not according to an occupation probability. Each bond is either open or closed. Open bonds allow the disease to be transmitted. Closed bonds do not. A disease starts at a source and flows along any open bonds, traveling from one susceptible person to another. It stops when it fails to encounter any new open bonds. The portion of the network that becomes infected from a randomly chosen starting point is called a cluster. Once a disease gets into a cluster, it will infect all the bonds within it.

Obviously, when the occupation probability is high and most bonds are open, the disease will spread to fill up most of the network. In this case, we see a **percolating cluster**, a single giant group of susceptible sites, connected by open bonds, that occupies the majority of the network. If the occupation probability is low and/or most bonds are closed, the disease will be contained and a percolating cluster will not form. In this situation, the groups of diseased individuals will be small and scattered.

The presence of a percolating cluster determines whether an epidemic turns into a pandemic. If a disease originates in the percolating cluster, all individuals within it will ultimately be infected. Under these conditions, it spreads very rapidly and is difficult to contain. If the disease starts outside the percolating cluster, it will eventually die out after having infected only a proportionately small number of individuals.

Percolation and Psychology

The percolation model gives us a theoretical perspective with which to understand the spread of activation in neural networks. If sites are neurons and bonds are synaptic connections, then we can determine how easily messages between neurons will spread. Some brain areas may have the equivalent of a percolating cluster. Neural activity that makes contact with the cluster will diffuse throughout it unimpeded. Several psychological conditions suggest the presence of percolating clusters in the human brain. These are seizures encountered in epilepsy, disorganized thinking in schizophrenic patients, and divergent thinking. We will now describe each of these in turn.

There are many different types of epilepsy, but they are all characterized by seizures. A **seizure** is an uncontrolled surge of synchronous electrical activity in the brain. For causes that are not well known, large populations of neurons in an epileptic's brain begin to fire simultaneously. In some cases, the seizure originates in one part of the brain, such as the temporal lobe or hippocampus, and spreads to other areas. Seizures result when there is too much activity in excitatory neurons and/or too little activity in inhibitory ones. Medications used to treat epilepsy work by decreasing activity in excitatory glutamate synapses. Others work by increasing the efficacy of inhibitory neurotransmitters such as GABA (gamma-aminobutyric acid). The neurons participating in a seizure are analogous to the sites in a percolating cluster. Once these sites are triggered, activation spreads rapidly through the entire cluster, much the same way a disease would spread throughout a population.

One of the symptoms of schizophrenia is **disorganized thinking**. It is characterized by a loose association between ideas and manifests itself in bizarre speech or "word salad." Susan Sheehan (1982) recorded what Maxine, a young female schizophrenic patient, said to herself one day:

> This morning, when I was at Hillside [Hospital], I was making a movie. I was surrounded by movie stars. The X-ray technician was Peter Lawford. The security guard was Don Knotts. That Indian doctor in Building 40 was Lou Costello. I'm Mary Poppins. Is this room painted blue to get me upset? My grandmother died four weeks after my eighteenth birthday. (p. 25)

The train of thought in disorganized thinking is illogical and jumps from one unrelated topic to another. It is as if the patient's brain state enters a neural percolating cluster that "takes over" their thought processes, causing them to leap along open bonds between disparate sites. Each topic that comes to mind in disorganized speech might then correspond to a localized area of neural tissue made up of a cluster of sites. The jumps in thought from one topic to the next seem to be the long-distance bonds linking these clusters together.

Creative individuals tend to score high on tests measuring **divergent thinking**. The goal of divergent thinking is to generate many different ideas about a topic in a short period of time. It typically occurs in a spontaneous, free-flowing manner, such that the ideas are generated in a random, unorganized fashion. One question measuring divergent thinking on a creativity test would be "How many uses can you find for a brick?" A highly creative person would produce more responses—such as "paperweight," "bookend," "sidewalk writing tool," and so on—than a less creative individual. Creative people are usually also better at generating ideas during **free association**, which involves stating whatever thoughts come to mind.

Divergent thinking seems to be a milder version of disorganized thinking, since it is controlled and more focused. During divergent thinking or free association, a person is allowing neural activation to flow along without inhibition. Ideas within a mental percolating cluster that aren't normally linked or are only weakly linked may then become related. In fact, when Evangelia Chrysikou and her colleagues (2013) used direct current electrical stimulation to slightly deactivate the left prefrontal cortex (a brain region

associated with cognitive control and inhibition), their participants scored higher in a divergent thinking test. Attention seems to be implicated in both disorganized and divergent thinking. During normal thinking, we may consciously and voluntarily direct our flow of thought to stay on topic. When this attentional focus is released, it may become easier to enter a percolating cluster of ideas.

The Future of Network Science

Network science is a powerful framework that we can use to better study cognitive phenomena. Researchers in this approach have uncovered basic architectural and functional features that underlie many different systems—whether those are brains, transportation networks, or ecologies. The application of network principles to the study of the mind should yield some fascinating insights over the next few years.

But we still have a long way to go. Many current network models assume that all nodes in the network are identical. This is clearly not the case. In many systems, the functioning of nodes can differ dramatically. Surely not every individual in a society thinks or acts alike, and we may discover that not every neuron operates identically either. The same signal sent to two nodes will not always produce the same response if the nature of the nodes differs.

We know that the brain is hierarchical, with processes operating at many different scales, from the synaptic to the cortical. This could pose a problem for network theory, where much of the dynamics occurs at a single scale. One way of dealing with this is to model hierarchical network processes. We could, for instance, place little small-world architectures inside bigger ones. We could then vary the type of connectivity between these networks at different spatial scales. Alternatively, we could implement networks that operate at different time scales—some running slowly, others more quickly. The application of network science to cognitive science has just begun and promises a number of fruitful approaches to the study of mind.

OVERALL EVALUATION OF THE NETWORK APPROACH

The network approach in general has received its share of criticism. Attacks leveled against connectionism are that its architectures are too simple and that its learning rules are inadequate. Semantic and propositional networks have their own problems related to organization and the proper role of inhibition. Having said all this, the network approach is still a strong perspective, both in terms of its theoretical assumptions and methodology. Connectionist networks are good at simulating learning and have some degree of biological plausibility.

The network approach poses a number of challenges to the classic information-processing view. Several differences need to be reconciled. These have to do with the numbers and sizes of processing units, and include the argument of serial processing versus parallel processing, as well as the argument of knowledge-based versus behavior-based approaches. What is needed, then, is a way of integrating the opposing views. New computers with parallel processing architectures are being developed. These machines may

yield new insights that will help bridge the gap between the classical and the connectionist perspectives. Another way to bridge this gap would be to have network researchers working with practitioners of other cognitive science disciplines in model building. Network models seem to constitute a "middle ground" between the actual biological structures studied in neuroscience and the more abstract representations of traditional cognitive science and artificial intelligence. Cooperative model building between these disciplines might resolve some of the discrepancies.

INTERDISCIPLINARY CROSSROADS: EMOTIONS AND NETWORKS

So far, we have been discussing how networks can serve to store cognitive representations such as concepts and propositions. Bower (1981) postulates that we can use networks to represent emotional states as well. For instance, we can take a mood such as "sadness" and assign it to a node in an associative network. When this node is activated, the person will then experience the corresponding mood. The sadness node through learning could have developed links not only to other mood nodes but also to nodes representing concepts and events. This would explain why a girl might feel sad when she thinks of an ex-boyfriend: The node standing for him was linked to the experience of being sad that happened during the breakup.

These "mixed" networks that combine emotional and cognitive representations can also explain why we might think of a concept when we are in a particular mood. In other words, the associative relation is a two-way street. Thinking of an ex-boyfriend might make a girl sad, but being sad might also make her think of her ex-boyfriend. This model can, therefore, predict the mood congruency effect where we are more likely to remember information that was learned in a particular mood if we try to recall it in that same mood later. We discuss this effect as well as other aspects of mood and memory in more depth in the emotional cognition chapter.

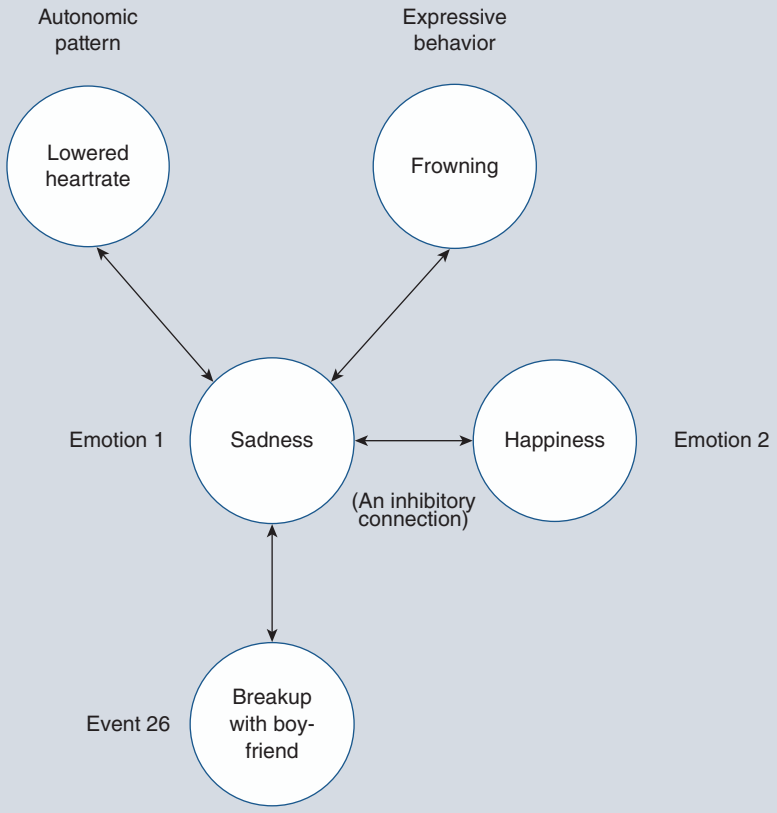
In Bower's (1981) model, there are six basic emotions, each with associated nodes that can be activated by various trigger factors. These factors become more developed over the course of the individual's life span. Presumably, some of these triggers are hardwired, as in the case of arachnophobia. They could also be acquired through experience, as when a person develops a learned association between a stimulus and a response. Figure 7.16 shows how a sadness node may be connected to others. Notice that there is an inhibitory connection between it and the happiness node, as these are opposing emotions. The happier you are, the less sad you will feel—and vice versa. In addition, if happiness inhibits sadness, not only will you feel less sad, but you will also be less likely to recall sadness-related information.

Several criticisms have been leveled at this emotional network model. For starters, not all moods elicit mood congruency effects. Anxiety does not produce the effect, while depression does. Second, people often distract themselves to either prolong positive states or avoid being in a depressive state (Rusting & Nolen-Hoeksema, 1998). For example, they can attend to mildly

(Continued)

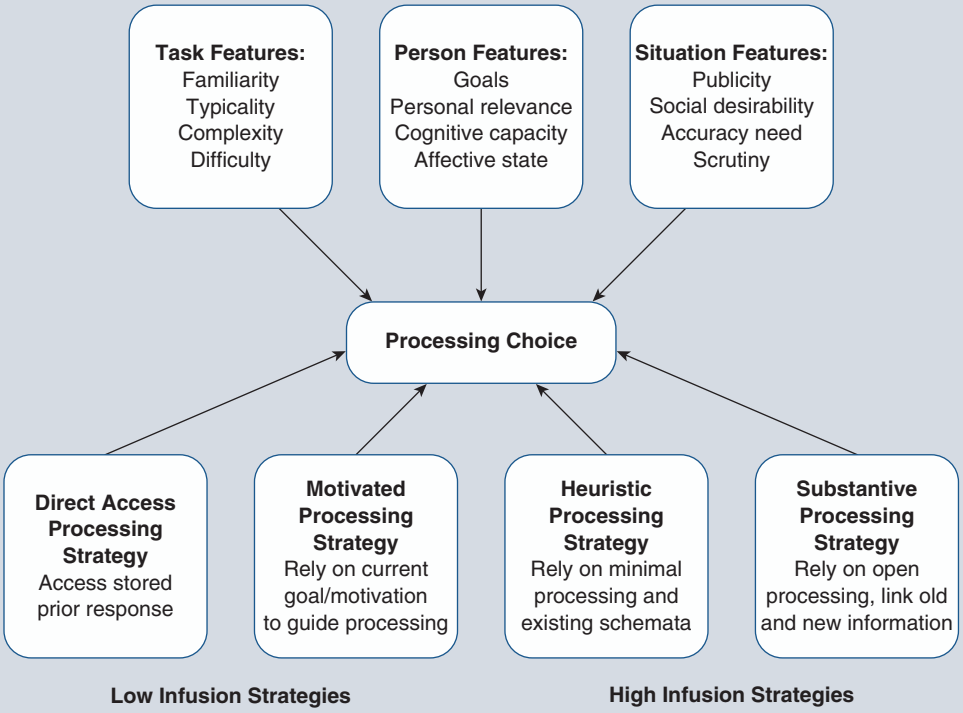
(Continued)

Figure 7.16 **An example of an emotion network. An emotion node can have links to other emotion nodes as well as to nodes representing types of physiological arousal, facial expression or other behaviors, and stored memories of related events.**



positive events, such as a comedy TV show, or actively think about uplifting events. Active distraction runs against the passive spreading activation that would induce same-mood cognitive biases. Rusting and DeHart (2000) induced a negative mood state in participants and then had them perform one of three tasks followed by memory recall. They were to ruminate about the cause of the negative mood, think of distracting thoughts, or simply free associate on whatever thoughts passed through their mind. The results varied depending on the condition.

Figure 7.17 The affect infusion model of emotion and cognition.



Source: Adapted from Forgas (1995).

Those who ruminated recalled more negative memories compared with the free-associating controls, and those who were distracted recalled more positive memories in comparison with the control group. The simple associative network cannot account for these results.

Forgas (1995) has developed the **affect infusion model** to explain the conditions under which emotions and moods infuse (influence) cognitive processing. These are depicted in Figure 7.17. When the situation involves a heuristic processing strategy, where one relies on minimal processing and the use of existing schema, then emotions produce mood congruency effects. Under these conditions, there are no simple responses available and the task is simple, familiar, and not of personal relevance. Emotional infusion also occurs when a substantive processing strategy is used. In this case, an individual would rely on an open processing

(Continued)

(Continued)

method that links old and new information. Under these conditions, the task is difficult, complex, or novel and there is no motivational goal.

However, there are also two situations in which emotions are unlikely to infuse cognition. The first of these is when we use a direct access processing strategy. This is one where we access stored prior responses, when the task is familiar, and there are no cues for more elaborative processing. Emotion also plays little role when we engage in a motivated processing strategy. In this situation, we rely on current motivations to guide processing because there is a specific motivational objective and there is a very specific search for information in memory that is required.

SUMMING UP: A REVIEW OF CHAPTER 7

1. In serial processing, computation is performed one step at a time. Contemporary computers operate this way. The brain, however, computes in a massively parallel and distributed way, with different ongoing operations at any given time and often using recurrent feedback signals.
2. ANNs simulate the way real neurons compute. They consist of nodes connected by links. Links have associated weights that can be positive or negative. A node's output is determined by multiplying its activation value times its weight. Multiple outputs typically converge on a single node. If these values exceed a threshold, the receiving node "fires," producing its own activation.
3. The perceptron was one of the earliest ANNs. It could recognize simple visual patterns and was limited because it contained only two layers. More recent ANNs have an input, a hidden, and an output layer and are capable of more complex computation. In backpropagation networks, the pattern of activation produced by a stimulus is matched to a known response. The difference between the two is fed back into the net so that it gradually becomes capable of recognizing and classifying a broad array of stimuli.
4. Convergent networks show a large amount of activity that gradually diminishes with time. Oscillatory nets show regular and periodic activity. Chaotic networks demonstrate a mixture of regular and nonperiodic action.
5. Connectionism is the approach that favors the use of network models to understand cognitive processes. The advantage to this approach is biological plausibility, learning, and resistance to damage. A disadvantage is that it omits some obvious features of real-world brains and can lose data when new information is acquired.
6. In semantic networks, nodes stand for concepts. Activation in these networks can explain a variety of results from memory experiments, including how long it takes to verify sentences.
7. Network science is a relatively new field of endeavor. Investigators in this discipline study the architecture and function of all networks that show some common features.

8. Small-world networks are those whose nodes are separated by at most a few links. There are a mixture of both local "neighborhood" links and more global long-distance links. Information in these nets can spread very rapidly. Mostly, highly connected hubs form the connections in hub-based or aristocratic networks. Hubs are nodes with a large number of links.
9. Anatomical studies show that mammalian brains have a small-world network organization. This helps explain a wide variety of brain behaviors such as epilepsy, plasticity, and synchrony.
10. Percolation is the process by which information propagates through a network. In the disease model of percolation, infection spreads through open bonds and susceptible nodes. A percolating cluster refers to that component of a network through which all the sites are infected. Examples of percolation in psychology include seizures, disorganized thinking, and divergent thinking.

SUGGESTED READINGS

Anderson, J. (2000). *Learning and memory: An integrated approach* (2nd ed.). New York, NY: Wiley.

Buchanan, M. (2002). *Small worlds and the groundbreaking theory of networks.* New York, NY: W. W. Norton.

Gurney, K. (1997). *An introduction to neural networks.* London, England: UCL Press.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (Eds.). (1986). *Parallel distributed processing: Foundations* (Vol. 1). Cambridge: MIT Press.

Strogatz, S. (2003). *Sync: How order emerges from chaos in the universe, nature, and daily life.* New York, NY: Hyperion.

Watts, D. J. (2003). *Six degrees: The science of a connected age.* New York, NY: W. W. Norton.

CHAPTER EIGHT

THE EVOLUTIONARY APPROACH

Change Over Time

THE EVOLUTIONARY VIEW

Cognition doesn't come out of "nowhere." Natural forces create it. Animal and human cognition are products of these forces. If we wish to understand cognitive ability, then we need to understand how they work. These forces go by the name of evolution, a process that involves adaptation to changing environments. Animals, including ourselves, have minds that allow us to think and survive in a given environmental situation or niche. We have evolved capacities that allow us to perceive, pay attention to, remember, and problem solve so that we can acquire food, avoid predators, and mate and reproduce. All of our cognitive abilities are direct or indirect products of these earlier environments and have evolved in response to such physical and social demands.

In this chapter, we will examine the evolutionary approach and its relation to cognitive science from several perspectives. The first perspective is **comparative cognition**, where researchers examine the similarities and differences in cognitive ability between different animal species. In this section, we will see how different environments have shaped different cognitive skills in animals. Birds that need to store large numbers of seeds and nuts, for instance, have evolved special memory and search capacities to find their stored caches of food. Animals that don't have such needs may not develop such special memory and search skills. So this section will show us how specific needs in particular environments give rise to particular forms of cognition.

The second perspective, **evolutionary psychology (EP)**, is concerned with how the human mind came into existence. It attempts to describe the selection forces in our ancestral past that gave rise to our mental structures. Here, we will extend what we have learned in the comparative cognition section to people, showing which specific demands have driven different human cognitive skills. For example, we show that our ability to reason logically

Learning Objectives

After reading this chapter, you will be able to:

1. Describe the process of natural selection.
2. List some cognitive skills that are adaptive in certain environmental niches.
3. Discuss the role of brain–body size ratio in comparative cognition.
4. List the six properties of an evolved psychological mechanism.
5. Explain the base-rate fallacy in judgment under uncertainty.
6. Define *loss aversion* in behavioral economics.
7. Discuss the evidence on sex differences in cognition.

seems to fail when a problem is described in abstract terms, but it is quite robust when couched in terms of a realistic social situation. Such results suggest that our reasoning capacity did not evolve in response to scoring higher on the SAT but in response to detecting cheaters in a group.

A field that is related to EP is **behavioral economics**. It focuses on how evolution may have shaped the way we make decisions involving money, debt, and other resources. Behavioral economists apply evolutionary principles to the domain of economics. Humans are economic creatures. For many thousands of years, modern humans have lived in societies where we must produce, consume, and exchange goods and services. Living in these societies affects our reasoning with regard to money. Take, for instance, the situation where someone else is given money but only offers you a small amount. Most people under such circumstances refuse to accept it. This is an irrational decision as it is free money, but it makes sense in a society where people must share reciprocally to survive. The rejection is a statement of distrust and could even get the greedy lender kicked out of his group for failure to cooperate. In conclusion, we will wrap up the chapter with a brief discussion of sex differences in cognition and how evolutionary forces might or might not account for them.

Other cognitive science approaches, in addition to those just mentioned, have adopted and used evolutionary principles. In the psychology chapter, we saw the functionalists asking the same questions that evolutionary psychologists ask. These functionalists wondered why we have certain mental abilities, for what purpose a given ability exists, and how a specific mental ability may have arisen. Evolutionary principles are also used in contemporary robotics, in which generations of robots are bred to produce robot offspring that are best adapted to the tasks that they have to carry out.

A number of important themes are touched on in this chapter. Evolutionary psychologists argue that many of our cognitive mechanisms are innate. This argument echoes the nativist position that we are born with knowledge. Evolutionary psychologists also argue that the mind consists of a set of special-purpose devices or **evolved psychological mechanisms** that are activated only within specific environmental contexts. This runs counter to the tenet of artificial intelligence and cognition that states that the mind is a general-purpose computing device, capable of solving virtually all kinds of problems equally well.

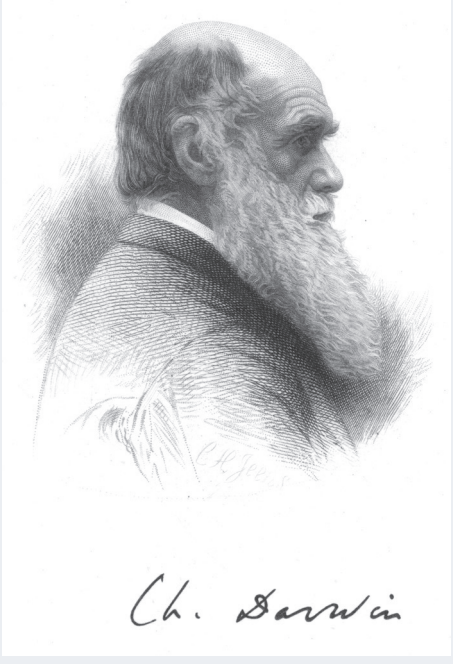
A LITTLE BACKGROUND: NATURAL SELECTION AND GENETICS

Before delving specifically into the evolution of mental structures, it is helpful to review some basic concepts of biological evolution. The biologist Charles Darwin (Figure 8.1)—in his classic 1859 work, *On the Origin of Species*—outlines the basic principles of his **theory of natural selection**, which is a description of the process by which animal species change over time. Darwin noticed that animals vary (within a species as well as from species to species) in all sorts of ways—for example, they have longer or shorter legs, bigger or smaller ears, or different coloration. The fact that animals differ in their physical traits is known as **variation**. Second, parent organisms pass on some of their gene-based characteristics to their offspring. This process is called **inheritance**.

A change in environmental conditions can make it more likely that animals with a specific trait will survive long enough to mate. Those without this trait might be more likely to die before mating. This process, whereby a particular attribute or attributes promote survival under altered conditions, is known as **selection**. The overall process by which animal species change in response to their environment is called **adaptation**.

Variation, inheritance, and selection are the vital ingredients of species change. Looking at the fossil record, we can see that there had been alterations in the foot structure of horselike animals. Early progenitors of the horse had feet with several toes. Figure 8.2 shows that, over time, these toes became fused, and the hooves of modern-day horses came into being. Originally, the members of this progenitor species varied with respect to toe structure. Some animals had toes that were closer to each other and less distinct anatomically. These animals were perhaps better able to run in an emerging grassland environment, as their feet were likely to have given them better support and locomotive power on flat terrain. Presumably, these horses ran faster and so were able to escape predators more easily. They were more likely to survive and pass on their adaptive trait to future generations.

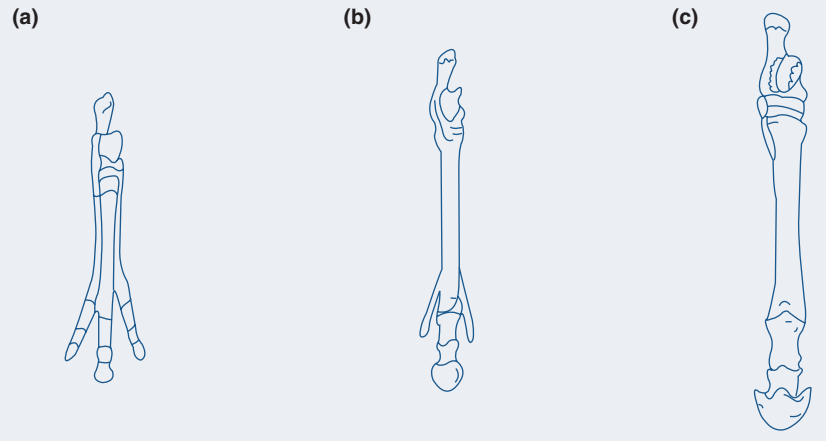
Figure 8.1 The biologist Charles Darwin (1809–1882) proposed the theory of natural selection.



Source: Photos.com/Thinkstock.

Natural selection is concerned with survival in a strict sense—that is, whether an organism simply lives long enough to reproduce, or not. Animals that possess an adaptive trait are better equipped to deal with a changed or changing environment. They are better at escaping predators, locating food, and keeping warm or cool, and so they are likely to survive long enough to reproduce. **Sexual selection**, in contrast, refers to the adaptation of traits that improve the organism's ability to attract a mate and, thereby, produce offspring. Some traits that are selected for naturally are not always selected for sexually. Females of some beetle species consider the presence of large horns on males of the species to be attractive. The males with these large horns mate more often and pass this trait to future generations more often. However, very large horns are also an impediment to survival. If they are too big, the males cannot easily escape predators such as birds. In this case, the optimum horn size represents a compromise between natural and sexual selection trading off against each other.

Figure 8.2 Evidence from the fossil record showing the evolution of the hoof from a toed foot in horse species.



COMPARATIVE COGNITION

In this section, we will discuss evolution in broad terms of how it has affected cognition in nonhuman animals of different species. This will serve as an introduction to thinking about the relationship between evolution and cognition and will set the groundwork for our later discussion, where we will focus on human cognition only.

Cognitive Adaptation in Animals

When looking at animals, we see that they are each adapted to a specific environment, what is called their **ecological niche**. Some birds migrate over vast distances and so need physical structures such as wings to enable them to fly, but they also need specific cognitive capacities that enable them to fly effectively. Studies show that carrier pigeons form an odor map of the environment that enables them to fly back more than 100 miles to a starting point after being released (Able, 1996; Wallraff, 1980). Pigeons may also be able to sense Earth's magnetic field to navigate, in effect using a built-in "compass" to steer them in the right direction (Wiltschko, 1996).

Some fish species, amphibians, and few aquatic mammal species have evolved the capacity to perceive electric fields. This sense enables them to detect prey that would normally be difficult to find using more typical senses, such as vision. Weak electric fields are given off by the muscles in an animal's body and can change when the animal moves, revealing its location. In most cases, this electroreception is passive, but the African elephantfish also has an organ in its tail that *produces* a weak electric field. Using this organ, the fish can send out an electrical signal that is then reflected back to the fish by objects in its immediate surroundings. Sensors on both sides of the fish's body pick up this electrical echo and allow it to estimate the distance of prey it can eat or obstructions in the dark (Von der Emde, Schwarz, Gomez, Budelli, & Grant, 1998).

Apes, monkeys, cats, dogs, and some birds can understand that an object, once it disappears from view, still exists. This ability is called **object permanence** and develops in human children at the age of 12 to 18 months (Piaget, 1938/1952). Object permanence is a primitive form of concept formation, because the animal must retain the idea that the object is still present in the absence of perceptual evidence. Gagnon and Dore (1994) found that dogs successfully look for a favorite toy that is hidden behind one of three boxes. In the natural world, having object permanence is crucial in locating animals that have disappeared behind obstacles such as rocks or trees. Even if a prey animal is not at the location where it disappeared, this is still a good starting point to continue the search.

Memory is a crucial adaptive capacity. It enables animals to remember where they have looked for something so that they don't have to expend valuable time and energy revisiting locations already searched. The Hawaiian nectar-feeding honeycreeper can remember which flowers it has collected nectar from and will avoid going back to those flowers if it has recently fed from them (Kamil, 1978). Memory also allows animals to recall where they have stored food. The Clark's nutcracker bird can store 33,000 pine nuts at 6,000 different sites during the summer and remember all these locations when it comes back to eat them in the winter and spring! Many different animals have both short-term and long-term memories and demonstrate some of the same memory phenomena that people do, including primacy and recency effects and proactive and retroactive interference (Bolhuis & van Kampen, 1988).

In addition to memory, **foraging** skills are also something that humans and other animals have in common. Whether searching for food that they've stashed away themselves or foraging for new food, various animals exhibit a particular statistical pattern in how they search an environment for food. For instance, spider monkeys, penguins, turtles, and tuna all show the same statistical pattern of making many short

movements in a small region while foraging, then occasionally making one long movement to a distant region where search resumes with more short movements (Kello et al., 2010). This same statistical pattern, sometimes called a Lévy walk or a power law, also shows up when humans move their eyes across a photograph of a city street (Rhodes, Kello, & Kerster, 2014) or when they play a foraging video game (Kerster, Rhodes, & Kello, 2016) or even when they try to recall as many vegetable names as they can (Rhodes & Turvey, 2007). This Lévy walk pattern of foraging in an environment is a highly efficient search strategy that is even found in ancient fossil records of the movement of sea urchins (Sims et al., 2014). Whether you are recalling category names from memory, looking at a complex scene, or hunting for your missing car keys, a style of foraging that involves *intensive* search in a narrow area followed by occasional *extensive* search across to a distant area appears to be evolutionarily conserved across humans and many other animals.

Nonhuman animals are also able to reason and solve problems as people do. Remember from the psychology chapter that Kohler's chimps demonstrated insight learning. They were able after some time to realize that a box and a stick could be used together to get at food. Another reasoning skill is **transitive inference**. This is equivalent to deductive reasoning, which we previously introduced. In a transitive inference problem, an animal must realize that if A is bigger than B and B is bigger than C, then A must be bigger than C. McGonigle and Chalmers (1977) presented squirrel monkeys with paired colored food containers, only one of which contained a hidden peanut. If the red container contained the peanut in one trial and the green did not, on the next trial, the green container would contain the peanut but the blue would not. Based on this training, the animals could eventually learn that the red container would have the peanut rather than the blue in a test trial.

Later research has shown that pigeons and rats can also solve transitive inference problems (Davis, 1992; von Fersen, Wynne, Delius, & Staddon, 1991). What is the evolutionary advantage of such a skill? According to Wynne (2001), it allows an animal to grasp rank orderings. This is particularly important in social animals with a hierarchical power structure. If Chimp A is dominant over Chimp B and Chimp B is dominant over Chimp C, one can infer that Chimp A is dominant over Chimp C without having to directly test this relationship, which could result in a fight and physical injury. Table 8.1 shows the cognitive skills of some other animal species and the adaptive selection pressures that gave rise to them.

In each of the cases outlined above, specific selection pressures have shaped particular cognitive skills. Note that these skills are often shared between members of different species. The key difference here is in the adaptive problem that needs to be solved, not the species in question. We would expect that all animals, regardless of their genetic and physical structure, might evolve similar cognitive capacities to solve similar problems. However, this is not an ironclad rule, as it is also the case that species can develop different solutions to similar problems; witness the different ways in which a bird's and bee's wings operate. In concluding this section, we need to mention that we do not address animal linguistic abilities here because those will be covered in the next chapter on language.

Table 8.1 Cognitive skills for different animal species and the adaptive function they serve.

Species	Cognitive Skill	Behavior	Adaptive Function	Reference
Lions	Counting	The greater the number of lions in a pride that are roaring, the less likely another pride will approach.	To avoid potentially harmful interaction with other competing groups	McComb, Packer, and Pusey (1994)
New Caledonian crows	Tool use	They nibble strips off the stiff edges of leaves so that they can stick them into holes to extract insects.	To obtain food	Bluff, Weir, and Rutz (2007)
Baboons	Deception	Subordinate males groom females who are hidden behind rocks out of view from dominant males.	To gain access to females for sexual reproduction	Byrne (1995)
Chimpanzees	Self-concept	Chimps who have colored marks applied to their faces while anesthetized will touch these marks while looking in a mirror.	Possible basis for intentionality and theory of mind	Gallup, Anderson, and Shilito (2002)

INTERDISCIPLINARY CROSSROADS: EVOLUTIONARY PROCESSES AND ARTIFICIAL LIFE

Artificial life (ALife) is the study of humanmade systems that have been designed to behave in ways that simulate the behavior of natural living systems (Langton, 1989). ALife researchers produce artificial "creatures" that inhabit virtual environments. They then study the behavior of

(Continued)

(Continued)

these creatures and how they interact with one another and with aspects of their environments. What makes ALife interesting and relevant to cognitive science is that although the rules that govern the creatures are often quite simple, the emergent behavior of these creatures is complex and can be considered intelligent. We discuss ALife in this chapter because these creatures learn and adapt via the use of evolutionary principles (e.g., Holland, 1992; Smaldino, 2014).

A good starting point for a discussion of ALife is Animat (Wilson, 1985). Animat is a single animal that lives in a simple environment. Its world consists of a square computer screen taken up by food items, which Animat can eat, and trees, which are obstacles Animat must navigate around. Animat can sense its environment; it is aware of the objects that are in its immediate vicinity. It can also act on its environment by moving one step at a time in one of the eight directions that are specified by the points of a compass (N, NE, E, SE, etc.).

Animat is initially endowed with a set of rules that govern its performance. These rules are in the form of classifiers, templates that match an environmental condition that Animat is subject to at a given moment in relation to an action. Classifiers are really a modified version of if–then production rules. Each classifier has an associated strength or fitness. Classifiers with greater strength values are more likely to be used by Animat, as they promote Animat's survival. Animat's actions also have consequences. A good action is one that promotes its survival, such as its moving to a location where there is food. Good actions are rewarded by increases in the strengths of their classifiers. Neutral actions, ones that have no effect on the well-being of Animat, such as bumping into a tree, engender no change in classifier strength.

Animat is "born" with a collection of classifiers. These classifiers, along with their strength values, are randomly determined and, as such, give Animat no inherent advantage. These starting classifiers may be considered equivalent to the genetic variation that exists among individuals or the members of a species and is shaped by selection forces. As Animat moves around its little world, classifiers are chosen and their strength values are altered on the basis of a system of reinforcement in such a way that survival-enhancing actions become more likely. However, this system alone is not enough to ensure Animat's survival. What Animat needs is a way of creating new classifiers that will better attune it to its world. This is where evolution comes in.

Wilson (1985) equipped Animat with an evolutionary algorithm. Two classifiers that correspond to the same general type of action are selected. Individual components of these classifiers are then swapped. Segments of each classifier are randomly picked and exchanged and then used to create new classifier offspring. This is, of course, analogous to sexual reproduction in which genetic material from both parents is used to create children. Wilson, however, also allowed for asexual reproduction. In this case, individual classifiers are cloned (reproduced in their entirety) to have random mutations.

The result of all this is that Animat's behavior becomes gradually more adaptive. Rather than modify the probabilities of its selections that represent its existing genetic endowment, it is able to generate entirely new actions, at least some of which will have beneficial results. It has been shown that these Animats, along with some more recent versions, learn quite rapidly. In one simulation, Animat was able to find food in only four steps compared with a chance-level performance in which it did the same in 41 steps.

Other ALife programs demonstrate greater complexity. Ackley and Littman (1992) have created an entire population of creatures. Some

are herbivores that eat plants, while others are carnivores that hunt and eat herbivores. Over time, the herbivores learned to "climb" trees to avoid predators, while the carnivores learned to cannibalize their fellow carnivores that died of starvation! Other emergent social behaviors of ALife creatures include parasitism, symbiosis, and flocking.

ALife simulations show that evolution-based software entities interacting in a diversified environment are capable of complex behavior. If we define intelligence as one's ability to modify one's behavior in an advantageous way over time in the pursuit of goals, then Animat and other ALife creatures may be labeled as intelligent. There is, however, no generally accepted definition of intelligence; therefore, the issue of whether such creatures possess it remains unresolved.

The field of ALife is closely tied to the study of artificial intelligence. One of the key differences between the two fields is that ALife attempts to model the behavior of the entire organism in an environment rather than just the functioning of the mind or some specific mental process (Beer, 1990). Because the behavior of ALife creatures is mostly reactive, their understanding of their world may be referred to as "minimally cognitive." However, a form of knowledge representation can still be identified in their internal processes (Beer & Williams, 2015).

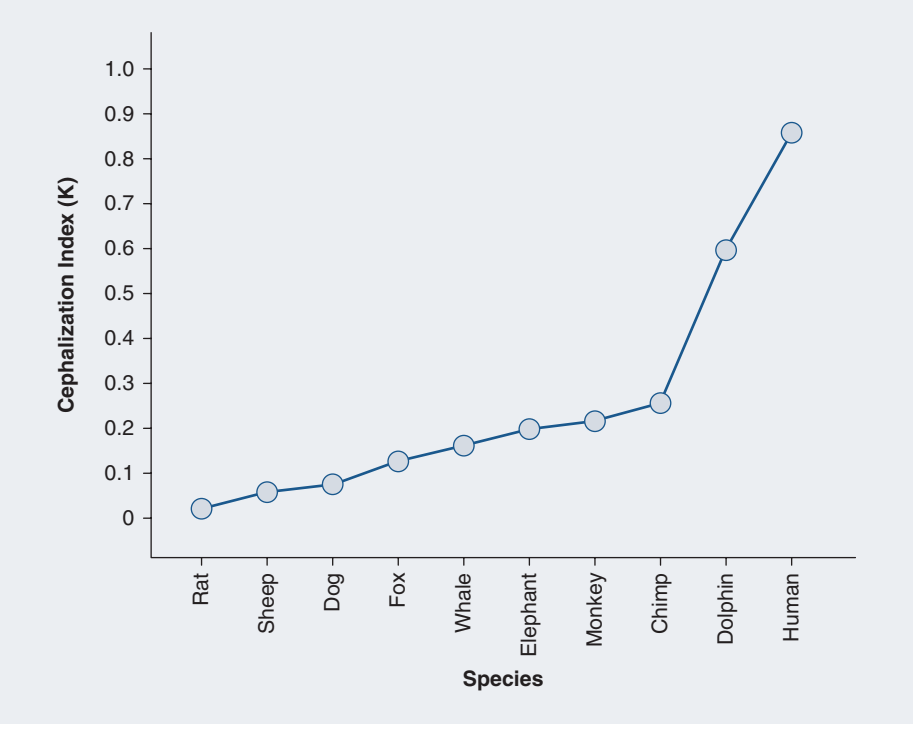
Comparative Neuroscience

A common question that may come to mind when thinking about cognitive differences between species is brain size. One might expect that "smarter" animals have bigger brains, or brains that differ qualitatively from less intelligent species. Researchers first examined the size issue by ranking animals in terms of their brain size. When this is done, we find that elephants and whales have the largest overall brains, in excess of 5 kilograms. Intuitively, this doesn't seem right since we humans assume we are the most intelligent species and we have smaller brains than those animals. Larger animals generally have larger brains because they have a greater body size. What we need, then, is a measure that expresses brain size in proportion to body size.

This measure is called the **cephalization index** (expressed as the letter *K*). High *K* values indicate that an animal has a larger-than-average brain for its body size, while a low *K* value indicates the opposite. Now, we see that humans are at the top of the list with *K* values of 0.89. In second place are dolphins with a value of 0.64. Primates and whales come in third with values of 0.2 to 0.3, whereas other mammals fall below 0.2. This is depicted in Figure 8.3.

However, we need to interpret this index with caution (Kaas, 2000). Other factors not related to intelligence influence brain size. Birds, for example, have evolved compact brains because extra weight interferes with flying. Dolphins, on the other hand, may have exceptionally large brains because weight is less of an issue in a buoyant aquatic environment. Another complication is that not all parts of brains are involved in cognition. The hindbrain in many animals does little more than regulate basic physiological functions, such as respiration, heart rate, digestion, and movement. So if we were to make our

Figure 8.3 The cephalization index allows us to compare brain size relative to body size across different species.



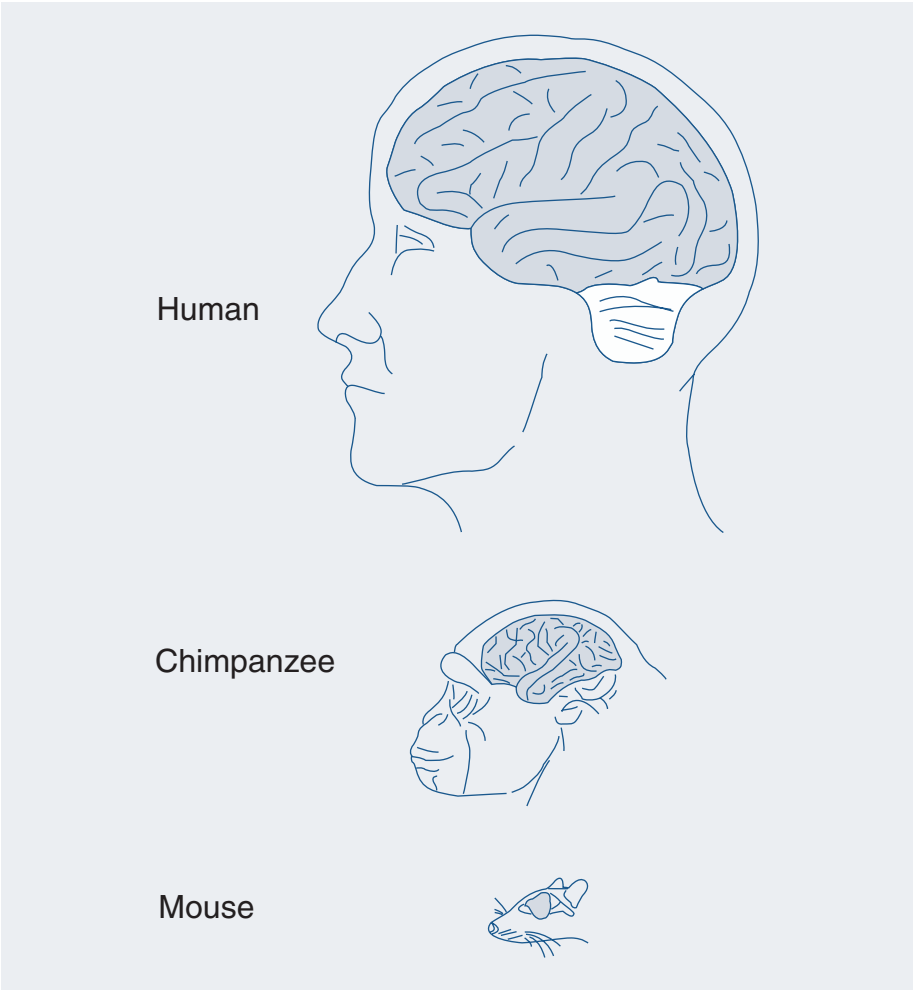
Source: Russell (1979).

comparison correctly, we would need to express cephalization only in terms of the cognitive brain areas rather than the entire brain.

The brain structure that underlies most cognition is the neocortex. In humans, the neocortex constitutes 80% of total brain volume. For primates, this is 50%; in rodents, 30%; and for insectivores (mammals that eat insects), only 13%. However, even this measure has problems. To start, only mammals have a neocortex. Other species, such as birds, do not possess one, so this prevents us from comparing neocortex size across all species. Figure 8.4 shows the brains of several different animals.

Not only can brain mass and neocortex mass provide insight into what makes certain species appear more intelligent, but the *neural connectivity* of those brains gives perspective on this as well. The neural *connectome* of humans shows more connections among nearby brain regions than is seen in monkey and cat brains (Van den Heuvel et al., 2016). That said, a mouse brain, with its large brain mass relative to body size, also has a rather dense neural connectome as well.

Figure 8.4 Brain size in several different animal species.



Another neural marker for intelligence may be **von Economo neurons**, also called spindle neurons. In humans and apes, these cells are found in the anterior cingulate cortex and the frontoinsular cortex. In people alone, they are also located in the dorsolateral prefrontal cortex (Fajardo et al., 2008). They are also found in beluga, humpback, killer, and sperm whales; bottlenose dolphins; and elephants. They allow for the fast transmission of signals across larger distances as would be necessary in animals with larger brains. These cells, in conjunction with a small-world network architecture, could promote rapid global information flow.

The location of von Economo cells in the cingulate cortex is no accident. The anterior cingulate cortex plays a vital role in higher cognitive function. It is involved in focused problem solving, the recognition of having made a mistake, and the initiation of correcting a mistake (Allman, Hakeem, Erwin, Nimchinsky, & Hof, 2001). Neurons in this region become active in experiments where participants must differentiate a correct response from conflicting cues and when there is a discrepancy between a desired state and a current state. It has been hypothesized to play a role in self-control and in the regulation of emotion and drive states, all signs of higher intelligence.

So what other measure can we use? One suggestion is the number of nerve cells. Unfortunately, the size and density of neurons in brains of different species differ. Some species, such as dolphins, have large cortical volumes but relatively few neurons. Other species, such as humans, have smaller volumes but more neurons. Although the comparison between dolphins and humans on this measure may correspond to actual intelligence differences, we cannot be sure that some other neuron property such as the number of dendritic connections is confounded with number. Future possibilities might include physiological function measures rather than anatomical differences. It may turn out that evolved intelligence also correlates with the rate of cortical development in the young animal (Finlay & Workman, 2013), the ease with which new synaptic connections are forged (Anderson & Finlay, 2014), or the complexity of neural circuitry (Van den Heuvel et al., 2016).

Evaluating the Comparative Approach

Two methodological issues plague the comparative approach (Wasserman & Zentall, 2006). Many of the studies investigating cross-species animal abilities have used different techniques. This makes them hard to compare, since any differences can be attributed to the different ways in which the experiments were conducted. A second problem is that some of the studies are done in artificial laboratory settings, while others are based on the naturalistic observation of animals in the wild. The laboratory studies can control for extraneous variables or alternate explanations of behavior but may be looking at actions the animals wouldn't normally engage in. The naturalistic observation studies yield natural behavior but have less control over other factors.

Perhaps one of the largest problems in this field is in how one attributes mental states to animals. People throughout history have had a tendency to attribute humanlike characteristics to animals, what is called **anthropomorphism**. Unfortunately, animal researchers are not immune to this tendency and sometimes ascribe more advanced cognitive abilities to animals than the data actually warrant. However, we must also guard against **anthropodenial**, which is a blindness to the humanlike characteristics of animals, or the animal-like characteristics of ourselves (De Waal, 1997a). We should not assume that humans are categorically unique in the arena of animal cognition. After all, we share a large number of brain structures and behaviors with many of them (Finlay & Workman, 2014). According to this perspective, we are not so different from some animals, and it may be easiest and most accurate, at times, to describe their behavior in the same way that we would describe our own, in terms of internal intentions, desires, and beliefs. In fact, in much the same way that Howard Gardner (2011) suggested that there are

multiple types of intelligence (in Chapter 3), Frans De Waal (2016) suggests that multiple types of intelligence may be found in the evolutionary tree of various animal species. For example, some of the forms of social intelligence displayed by nonhuman primates may have evolved into our own human versions of empathy, morality, and justice.

However, we still have to be careful about potentially overinterpreting the behavior of nonhuman animals. Cheney and Seyfarth (1990) observed a low-ranking vervet monkey named Kitui utter a leopard alarm call whenever a new male tried to join his group. The addition of a new male would lower his status and introduce more competition for him for food and mates. The call caused all the monkeys to scatter into the trees and prevented the entry of the new male. At first blush, Kitui's action might make it seem as though he was capable of deceit and perhaps had a theory of mind, an understanding that the other members had beliefs. However, further study showed that he came down from the tree and uttered the cry again closer to the intruding male. By coming down from his hiding spot, Kitui would reveal to the others that he really didn't believe a leopard was there. This would negate any deceitful action and cause the other group members to suspect him. So it is perhaps unlikely, given this behavior, that he actually intended to deceive or had a developed theory of mind.

Other examples of overattributing sophisticated cognition to animals come from studies of imitative behavior. One reason an animal might imitate another is because it understands that the other animal is performing an action to achieve a desired goal, such as acquiring food. The animal in question would then copy this action to obtain the desired goal for itself. Studies of potato washing in Japanese macaques and following behavior in ducks have instead shown that imitation can be a consequence of **stimulus enhancement**, where attention simply has been drawn to the location or the behavior and there is no inference of desired goals. Many other abilities ranging from concept formation to self-perception can sometimes be accounted for by simpler explanations that don't require sophisticated thought processes. Not only is this true for nonhuman animals, but it is true for humans as well.

EVOLUTIONARY PSYCHOLOGY

We humans don't look like much. We are thin skinned—and almost gawky looking when we stand upright. To a passing predator, we must seem like a particularly easy kill. We have no armor, fangs, or claws with which to defend ourselves, nor are we particularly strong or fast in comparison with other animals. So what has enabled humans to survive and thrive in a harsh world? The answer, in a word, is our minds. The species *Homo sapiens* has the most sophisticated mental faculties of any species on Earth. Our capacities to think, to use language, to solve problems, and to strategically coordinate are greater than those of any other animal. An advanced mind allows for planning, communication, social cooperation, and the development of technology, which give us a tremendous competitive survival advantage.

But how did the human mind come into existence? What processes formed it? These are some of the questions asked by the evolutionary approach. In evolutionary theory, changes in an environment give rise to corresponding physical changes in an

animal species that better enable it to survive. Giraffes, to use an often-cited example of adaptation, have developed longer necks over many thousands of generations to better enable them to get at treetop vegetation. Similarly, we can consider the mind to be such a body structure—comparable to the giraffe's neck—that has been shaped by evolutionary processes. Evolutionary psychology (EP) not only attempts to describe these forces that are believed to have fashioned the human mind but also is concerned with understanding the mind's organization and function and how interaction with an environment produces observable behaviors.

Evolutionary psychologists face a fundamental problem. They need to understand the selection forces that gave rise to mental structures but cannot know precisely what these were because they correspond to events that were taking place far back in human prehistory. Evolutionary psychologists believe that the period of time during which many human psychological mechanisms evolved was the Pleistocene era, which began approximately 2 million years ago. This period is referred to as the **environment of evolutionary adaptation** (**EEA**; Bowlby, 1967). Because we cannot go back in time to study what occurred during this period, we must make inferences about the physical and social environment of that time based on evidence of various types. Buss (1999) describes the types of evidence used to test evolutionary hypotheses. These include archeological artifacts and records, data from studying hunter–gatherer societies, systematic observation of human behavior, and self-reports in the form of interviews and questionnaires. EP also relies on a variety of methods for the testing of its hypotheses. Among these are comparisons across species, comparisons of males and females within a species, and experimentation (Buss, 1999). The evolutionary approach thus adopts a wide variety of methods rather than relying exclusively on the scientific method.

Evolved Psychological Mechanisms

Evolutionary biology focuses on all aspects of an animal's makeup. It is concerned with how selection forces came to shape the individual parts of an organism and how these parts became integrated. The scope of EP is narrower—it examines the impact that these forces had on psychological mechanisms only. Whereas evolutionary biology looks at organs and organ systems, EP looks primarily at a single organ—the brain. This perspective views the mind as a collection of psychological mechanisms that have evolved and attempts to describe how each came to be and how each operates.

In the 1960s, the view in cognitive psychology was that the mind is a **general-purpose processor**. This processor can be "fed" a problem in any format and can come up with a solution by applying a known set of rules. Modern computers operate in this way. The advantage of this process is that it can solve all problems of a given kind, assuming that the correct values of the problem's variables are mapped onto the algorithm designed to solve the problem correctly. The evolutionary psychology perspective adopts a very different view. It argues that the mind is not general in any way but, instead, a set of many specific capacities. Each of these capacities is called an evolved psychological mechanism.

Cosmides and Tooby (1992) refer to the mind as a "Swiss army knife." In this analogy, each part of the knife unfolds to solve a different problem: The corkscrew takes out corks, the scissors cut paper, and so on. Likewise, each evolved psychological mechanism is assumed to have evolved in response to the need to solve a specific problem of adaptation and is activated by the contextual information of that problem alone. They are thus examples of modules we discussed previously. The general characteristics of modules are discussed early in Chapter 4. Table 8.2 lists the six properties of evolved psychological mechanisms and illustrates them with an example.

To sum up, EP views the mind not as a one-size-fits-all machine but as a collection of functionally distinct mechanisms. Evolutionary forces operating over vast periods of time created each of these mechanisms. Each evolved to solve a problem our ancestors faced during the EEA. It is important to keep this multiple-mechanism view in mind as we describe specific examples of cognitive processes. According to EP, for every possible cognitive process, it is not the brain as a whole that is acting (according to this perspective) but a particular component of the brain. Thus, when we talk about the evolution of memory or logic, we are talking about specific mechanisms that may constitute only a part of the brain's memory or logic systems.

Table 8.2 The six properties of an evolved psychological mechanism.

Property	Explanation	Example: Arachnophobia
1. Purpose	To solve a specific problem of survival or reproduction	Prevents being bitten by a poisonous spider
2. Perceptual focus	Mechanism triggered only by a visual image of a specific stimulus	Two body parts with eight legs
3. Communication	Input of mechanism informs the organism of the type of problem	Organism knows adaptive problem is one of death, not food selection or mating
4. Output	Input produces a behavioral output or response	Scream or swipe spider off body
5. Type of output	Can be physiological activity, input to another psychological mechanism, or a behavior	Heart acceleration and sweating, decision to swipe or run away, resulting behavior
6. Solution	Mechanism evolved to provide a solution to the problem	Removing spider from immediate vicinity reduces threat

Source: Adapted from Buss (1999).

Evolution and Cognitive Processes

In the following sections, we will describe specific evolutionary hypotheses that pertain to several domains in cognition and the evidence that supports them (Gaulin & McBurney, 2001; Palmer & Palmer, 2002). These hypotheses concern cognitive processes only, which include categorization, memory, logical reasoning, judgment under uncertainty, and language. The motivation that underlies this research is the desire to arrive at an evolutionary account of how these cognitive processes arose. In other words, why do we categorize and remember things the way we do? Why are we so poor at reasoning logically and making judgments under uncertain conditions? How is it that human beings alone of all the species on Earth have developed such a complex linguistic ability?

Categorization

Research conducted by Eleanor Rosch (1973, 1975) shows that people form mental categories not in an "either–or" fashion but, rather, in a continuously graded way. Why should this be the case? Why do we not lump things together into one group on the basis of a single characteristic—for example, "red objects" or "curved objects"? The organization of concepts, it turns out, is governed by "fuzzy" categories, whereby a particular item can be more or less representative of the category. In this section, we will examine why humans form concepts in this way. Recall the finding—from the chapter on networks—that a bird such as a robin is thought of as being more representative or prototypical of the category of "birds" than another type of bird—for example, a penguin. This is because robins tend to come closer to our idea of what a bird is—something that is small, sings, flies, and so on. Robins are thus more typical of "birds" than are penguins (see Figure 8.5). Participants in experiments respond more quickly when they are asked to make a true/false judgment about the statement "A robin is a bird" compared to the statement "A penguin is a bird"—the so-called **typicality effect** (Rips et al., 1973). In fact, if they are responding to the word *penguin* with a computer mouse, their mouse movement may curve slightly toward the "mammal" response option before it finally lands on the "bird" response option (Dale, Kehoe, & Spivey, 2007).

Categories of the sort found in nature are almost always continuous: They contain items that span the entire spectrum of typicality—from being very representative and prototypical to being quite unrepresentative and perhaps easily mistaken for items that belong to another category. The categories that result from human categorization processes, therefore, mirror the distribution of natural objects in the world. This suggests an evolutionary process whereby we came to acquire concepts and a sense of their relatedness on the basis of how the items that these concepts represented were ordered in the natural environment. Further evidence in support of this notion comes from Boster and D'Andrade (1989). They found that professional ornithologists as well as those who had little experience and/or knowledge of birds all grouped stuffed bird specimens in similar ways. Their groupings demonstrated typicality and hierarchical effects reflective of what exists in nature.

What purpose do typicality-based categories serve? They must have provided us with some sort of selective advantage—otherwise, they would not have arisen. The

Figure 8.5 Which of these birds looks more like a bird? What does this tell you about your concept of "birds"?



advantage is this: If we know something about an item we have experienced, then it becomes possible to form judgments that have to do with related items (Rips, 1975). Imagine encountering for the first time a plant with four leaves and purple spots. After eating this plant, you become ill. Several months later, you stumble across a plant with three leaves and reddish-purple spots. Would you eat it? Probably not. There is a reasonable chance that the two species of plant, because they are similar, share properties. Typicality categories thus allow us to capitalize on past experience. We generalize from what we know to what we don't know.

Memory

Other approaches in cognitive science tend to study basic memory processes. You will recall that cognitive psychologists are concerned with, among other things, how information is converted from short-term into long-term memory. Neuroscience can identify the brain structures that underlie this transformation. Connectionists can then create artificial networks to model the process. EP, on the other hand, cares less about memory function per se and more about the purposes our memories serve. EP researchers want to know why our memories are the way they are. Specifically, they want to know how the demands of our ancestral environment have shaped memory and have made our memories operate the way they do. In this section, we will examine one aspect of memory—the relationship between the frequency of contact with an item in one's environment and recall.

Every day, we are bombarded with a tremendous amount of information. Only a small fraction of it is actually remembered. This is a good thing, as most of the

information is trivial. For instance, can you remember the color of the shirt your best friend was wearing the previous Monday? It would make sense, then, for us to remember only the information that may be of some benefit to us and to forget that which doesn't benefit us. But what constitutes beneficial information? What things should we be better at remembering? An evolutionary account avers that information that we encounter more often should be better remembered. The more often we encounter an item, the greater its relevance to our daily life and our survival. And the less often we encounter an item, the lesser its relevance to our survival. Remembering your roommate's name is important—because he can help you in a variety of ways. You would also be likely to hear your roommate's name quite often, as you see him practically every day. Remembering the name of your niece who lives in another country is less important to you, as you are far removed from interacting with her on a daily basis. Correspondingly, you would very rarely hear her name, assuming that it is not the same as that of someone else you know better.

One way to test this hypothesis is to compare the frequency of our encounters with specific items with our ability to remember them. Clearly, this relationship will be positive: The more frequently we encounter items, the better our memory of them. In addition, the relationship is described by a power law function, similar to the statistical pattern seen with foraging skills in the comparative cognition section above. This specific relationship is represented when we plot on a log–log graph our ability to remember items versus how often we encounter them.

Anderson and Schooler (1991) tested this relationship. They examined several real-world sources of information and determined for each type of information the statistical function for the frequency of occurrence and recall. That is, they determined the mathematical relationship specifying how many times a piece of information occurred and how accurately it was remembered. They used newspaper headlines, utterances made by parents to their children, and names of the authors of e-mails that were received by one of the researchers over an approximate 4-year period. For all these sources, the probability that a particular utterance or word would crop up a second time was a decreasing function of the number of days since its previous appearance. For example, if the word *north* appeared in a newspaper headline yesterday, there would be a relatively good chance that it would appear again today, presumably because it was part of an ongoing news story. If *north* had been in the paper only in the previous week, the chances that it would appear again today would be much lower. The researchers then measured memory for these items and found that the function that described probability of recall over time was in accordance with the probability of occurrence. This shows that our memory for items is directly proportional to how often we encounter them in the environment. Our memories are better for items that we encounter more often and poorer for those we encounter less often. This finding supports the notion that our memories have evolved such that they retain items only to the extent that we need them. From an evolutionary point of view, this makes sense. We should remember those things that we encounter frequently, as it is very likely that they are of importance to us, and forget those things that we encounter infrequently, as it is likely that they are not of much use to our survival.

Logical Reasoning

Recall from the chapter on philosophy that deductive reasoning involves the application of logical rules to a set of premises. A conclusion is then derived: If the premises are true, then an application of the rules will always yield a correct conclusion. Deductive thought is thus a valid way of garnering new information about the world. Logic is used in other approaches to cognitive science too. Connectionists use the rules of logic to describe how neural units interact. Researchers in artificial intelligence create computer programs that use logic to generate new information from existing data and to solve problems. In the evolutionary approach, investigators want to know why we as humans are so bad at thinking logically and to relate this deficiency to the kinds of social conditions our early ancestors may have faced.

If you have ever taken a course in logic, you may have wondered why the subject is so difficult. The rules of logic are straightforward. All one has to do is memorize the forms of these rules and then plug into them the relevant aspects of the problem. In this fashion, one can then apply those forms to any problem and be assured of obtaining a correct solution. Unfortunately, it is not as easy as it sounds. Most people find it difficult to reason logically. Interestingly, however, there is an exception. People can reason logically under certain conditions. The evolutionary approach can tell us why we are so bad at logical reasoning in general and what the exception is.

The **Wason selection task** is designed to measure a person's logical thinking ability. Figure 8.6 gives an example of such a task. Take a minute now and try to solve this problem. You probably found that it was quite difficult. Next, try to solve a different version of the problem, shown in Figure 8.7. When you have finished—or if you are having trouble—consult Figure 8.8 for the correct solutions. If you are like most people, you would have found the "bouncer" version of the problem to be much easier. Why is this? Both versions have the same underlying logical structure and should be of equal difficulty.

Figure 8.6 **You have been hired as a clerk. Your job is to make sure that a set of documents is marked correctly, according to the following rule: "If the document has an E rating, then it must be marked code 4." You have been told that there are some errors in the way the documents have been coded and that you need to find the errors. Each document has a letter rating on one side and a numerical code on the other. Here are four documents. Which document(s) do you need to turn over to check for errors?**

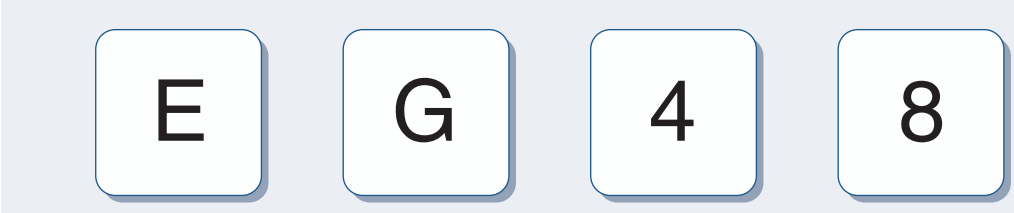


Figure 8.7 You have been hired as a bouncer in a bar, and you must enforce the following rule: "If a person is drinking vodka, then he or she must be over 20 years old." The cards depicted in the figure contain information about four people in the bar. One side of each card lists a person's age, and the other side specifies what he or she is drinking. Which card(s) do you need to turn over to make sure that no one is breaking the law?



Figure 8.8 Solutions to the "clerk" and "bouncer" problems.

Solution to the "clerk" problem.
Rule: E-rated documents must be marked code 4.
Step 1: Check E-rated documents to see if they meet the coding criterion.
Step 2: Check any documents not matching the coding criterion (e.g., code 8) to see if they are E-rated.
No more checking necessary:
G ratings (paired with any code) could not violate the rule.
Code 4 documents do not have to be E-rated.

Solution to the "bouncer" problem.
Rule: Vodka drinkers must be over 20 years old.
Step 1: Check vodka drinkers to see if they meet the age criterion.
Step 2: Check anyone who does not meet the age criterion (e.g., 17 years old) to see if they are drinking vodka.
No more checking necessary:
Pepsi drinking (at any age) could not violate the rule.
People over 20 years do not have to drink vodka.

According to Cosmides and Tooby (1992), the bouncer problem is easy because it involves **cheater detection**, the ability to discern who has taken as his own something he does not deserve. They argue that during the EEA, human societies were characterized by **reciprocal altruism**, the sharing of hard-won resources among group members. In

this kind of social cooperative, it is important to detect "freeloaders"—those who take more than their fair share. Cosmides and Tooby believe that this social environment shaped our reasoning abilities such that when we are presented with a logic problem that involves cheating, we can solve it. If a logic puzzle possesses the element of a cheater, then the evolved psychological mechanism for cheater detection is activated. This mechanism "understands" the rules of logic and can apply them, but only in this specific context. Note that this understanding stands in stark contrast to that which maintains that the mind is a general-purpose processor that can solve logic problems of any sort.

There is abundant experimental evidence in support of the cheater detection model of reasoning. One study comes from Gigerenzer and Hug (1992). They presented participants with two Wason selection tasks, both of which entailed a social contract of the following sort: "If you stay overnight in a mountain shelter, you must help out by carrying up some firewood to add to the supply." Both of these tasks were roughly equivalent with respect to content and logical structure. The only real difference between the two was that one of the tasks involved cheating, while the other did not. The results showed that 78% to 90% of the participants were able to solve the cheating version. Only about half as many participants were able to solve the noncheating version. This shows that the framing of a logic problem as a social contract alone is not enough to elicit the problem-solving skill; it is elicited by a social contract that specifically involves cheating. In another study, it was discovered that even children a little older than 2 years of age seem to have this cheater-detecting capacity (Cummins, 1999).

Judgment Under Uncertainty

In many situations in life, we have to make a decision in the absence of complete information. What are the chances that Mary will go out with me? What is the likelihood that I will make the college swim team? As is the case with logical reasoning, most people are quite poor at solving problems of this type—problems that involve probability (Nisbett & Ross, 1980; Tversky & Kahneman, 1974). Why is this? One reason is that humans facing these kinds of problems rely on heuristics. A heuristic is a mental "rule of thumb" or strategy that is a fast and easy way of solving a problem. The problem with heuristics is that they aren't always right. Heuristics can lead us to commit **fallacies**—fundamental misunderstandings of statistical rules. The role that evolution has played in shaping our thinking in this situation is discussed below.

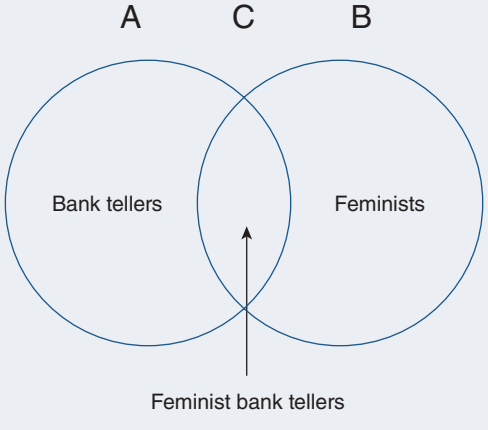
Tversky and Kahneman (1974) were the first to study heuristics and fallacies. In one study, they gave the participants a description of a person. Participants were also told that this person belonged to a set of 100 people. There were two conditions in the experiment. In the first, the participants were told that the set consisted of 30 lawyers and 70 engineers. In the second, they were told that the set contained 70 lawyers and 30 engineers. Here is a sample description:

> Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies, including home carpentry, sailing, and mathematical puzzles.

Is Jack a lawyer or an engineer? The participants answered overwhelmingly that Jack was an engineer, even in the condition in which engineers made up only 30% of the set. Tversky and Kahneman (1974) argue that this kind of error is due to a **representativeness heuristic**, the tendency to judge an item on the basis of its perceived similarity to other items. Because the description of Jack fits the stereotype of the engineer, we consider it very likely that he is one. The representativeness heuristic in this case is accompanied by the **base-rate fallacy**, ignorance of the base rates that define the set of 100 people.

Here is another example of our difficulty in making judgments under uncertainty (Tversky & Kahneman, 1974). Consider the following description: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. In college, she was involved in several social issues, including the environment, the peace campaign, and the antinuclear campaign." Which of these statements do you think is more likely: "Linda is a bank teller" or "Linda is a bank teller and is active in the feminist movement"? Most people overwhelmingly choose the latter sentence, even though it cannot possibly be the more likely of the two, since there are always going to be more tellers than feminist tellers (see Figure 8.9). Participants again rely on the representativeness of Linda's description, which fits the stereotype of the feminist. In this instance, participants ignore the conjunction rule, which states that the probability of encountering those who are both feminists and tellers is lower than the probability of encountering those who are one or the other. The making of this error is known as the **conjunction fallacy**.

Figure 8.9 **A Venn diagram illustrating the conjunction rule. Circle A represents all bank tellers. Circle B represents all feminists. Region C, the portion of overlap between the two circles, represents feminist bank tellers and will always be smaller than the regions that correspond to the number of bank tellers or feminists, considered separately.**



So is it fair to say that humans cannot think in terms of probability or uncertainty? From an evolutionary standpoint, one would guess that this is not a fair statement, as uncertainty is an inherent characteristic of the natural world. Gigerenzer and Hoffrage (1995) argue that the difficulties that human beings have in this area, which are documented in the literature, stem not from a fundamental inability on our part but from the way that these types of problems are usually structured. Many of these problems entail probabilities that are expressed as fractions or percentages rather than as frequencies, expressed as numbers of occurrences. Probabilities are more complex numerically, as they require dividing one number by another. They also are a rather recent invention in mathematics, just a few hundred years ago. By contrast, modern humans have been encountering frequency data as part of their daily existence for many thousands of years. We modern humans can easily think about how many times we've eaten today or how many rabbits we caught yesterday or how many games the New York Yankees have won this season.

It makes sense, then, that the psychological mechanism that underlies uncertain reasoning should be sensitive to frequency data. It turns out that this is, indeed, the case. Gigerenzer and Hoffrage (1995) assessed a number of uncertainty problems and converted them from a percentage to a frequency format. They found that roughly three times as many participants answered correctly. Their finding, similar to the findings of the logical reasoning study illustrated earlier, is another example of how problem solving is narrowly tuned to a particular type of information. This attunement to a particular type of information is the second major characteristic of evolved psychological mechanisms and again runs contrary to the notion of the mind as a general-purpose processor—at least within these problem-solving domains.

Yet another example of poor probabilistic reasoning is the **gambler's fallacy**. Do you think that someone who has been losing at a slot machine for hours is more likely to have a jackpot coming up soon than a person who has been losing at one for only 10 minutes? Do you think that a fair coin that has come up "heads" eight times in a row is more likely than usual to come up "tails" on the next toss? If you answered yes to either of these questions, then you have been taken in by the gambler's fallacy. It is the belief that the probability of an event is influenced by how often it has occurred in the past—even when those events are independent of one another. If a particular outcome has turned up quite often in the recent past with a random chance event, people are apt to believe that there is an increased likelihood that a different outcome will occur next. In reality, the probability of each event in these gambling scenarios is independent. The chances of winning at slots the 100th time are exactly the same as for the first. The chance of getting a "heads" in a fair coin flip is 0.50 on every flip.

Why, then, do we persist in holding such beliefs? One possible explanation is that independent outcomes are typically the product of artificially engineered devices that are not encountered in the natural world, such as dice or roulette wheels (Pinker, 1997). In nature, the past often *is* the best predictor of the future. If it was sunny yesterday, there is a good chance that it will be sunny today. If a herd of mammoth migrated through a particular region the previous year at this time, the chances are that they will do it again this year. Natural events tend to have regular time courses and shared causes; therefore, some knowledge about the way an event unfolded in the past is informative about the

way similar events will unfold in the future. One might think that this evolutionary experience would cause the gambler's fallacy to go in the opposite direction (i.e., a person who sees a coin flip result in "heads" eight times in a row would conclude that the next flip is *more* likely to be "heads" again). However, humans are accustomed to alternating cycles between opposing states, such as night following day or a half-moon following a full moon. Therefore, in considering opposing possibilities and taking the past into account, we sometimes fall prey to the gambler's fallacy.

Language

Language is so important in cognitive science that it is a separate field of study. Linguists study the organizational rules of language and how language is acquired. They also construct models of language processing. From an evolutionary standpoint, the focus again is not so much on processing but on how language came to be. Because language possesses a social function, evolutionary researchers want to know what aspects of early human social life might have given rise to its development.

One of the most obvious things that sets *H. sapiens* apart from other species is language. Other animals can communicate, and primates can be taught rudimentary linguistic skills, but nowhere else in the animal kingdom do we see such a sophisticated use of and widespread reliance on language as in the human species. The biological and cultural evolution of language is believed to have catalyzed the development of intelligence, social cooperation, and other phenomena, in the process of which human beings derived great benefit. Researchers in this area have argued that language is an evolved trait and have provided several theoretical explanations for how it could have arisen. Nonevolutionary accounts of language are discussed in Chapter 9.

Converging lines of evidence point to the possibility that human language ability has evolved (Palmer & Palmer, 2002). First, languages share a number of universal commonalities (Chomsky, 1986; Hockett, 1960). Among these are meaningfulness, the ability of those who possess language to communicate about things not physically present, and the ability of those who possess it to produce an infinite number of new meanings via the combination of existing elements. Almost all human babies progress through the same developmental language stages, in the same order, and on roughly the same time schedule irrespective of which native language they are learning. Children also seem to acquire the rules of language faster than they could possibly acquire them through formal learning. Finally, there are specific brain areas devoted to language. Most individuals are left hemisphere dominant with respect to language and have localized sites within this hemisphere that are devoted to language comprehension and language production. These observations suggest that language capacity may be prespecified and is present at birth. This, in turn, would mean that language capacity is genetically coded and has been shaped by selection forces that were operational at some time in our ancestral past.

Robin Dunbar (1996) has developed a fascinating theory of language evolution. He notes that in hominids, there is a positive correlation between group size and the size of the neocortex. The neocortex underlies cognitive function and is quite large in humans. This brain region processes social information—for example, information that has to do

with which members of a group are our friends and which members are our enemies. These kinds of social computations become more complex with increases in group size.

Most primates form alliances by grooming one another—that is, by picking parasites and dirt from one another's fur. This feels good to them and is a form of social bonding. But grooming, because it is a one-on-one activity, imposes a limit on how many individuals in a group can bond. Once a group becomes large enough, not every member of the group can be groomed. Language is also a form of social bonding and a way for individuals who possess this capacity to form alliances within groups. These alliances are forged primarily through the equivalent of gossip, wherein small groups communicate in both positive and negative ways about other group members. Because one can speak comfortably to a larger number of other individuals at once, language may have driven the formation of larger groups in early protohuman societies.

The coming into being of larger groups prompted further innovations, which were adaptive in nature. Larger groups are better able to defend themselves against predators in an open savanna environment. Language allows for complex, coordinated social behavior. Language would have facilitated hunting, foraging, child care, and other crucial survival behaviors. The development of language may have received an additional boost from sexual selection. Assuming that females recognize verbal skill as a rough indicator of intelligence, they might choose to mate with males who are good "talkers," as their intelligence would make them better partners.

Behavioral Economics: How We Think About Profit and Loss

The discipline of behavioral economics is devoted to studying the basis of economic decision making. It applies scientific methods to understand the cognitive, emotional, and social influences on the way we think about things such as money, goods, services, buying, and selling. The classical view in economic theory is that people are rational decision makers. This implies that we reason logically when presented with choices, always picking outcomes that maximize our benefits and minimize our losses. In reality, it turns out that we are far from rational when it comes to making economic choices, and our motives in these situations often have evolutionary explanations. As was the case with judgment under uncertainty, our behavior stems from various heuristics. Framing, or the way a problem is presented, also has a powerful impact under these circumstances.

Imagine that you are given $1,000 and that you can split it with somebody any way you want. This scenario in the lab is known as the **ultimatum game**. Being slightly selfish, you might decide to keep $900 and give $100 to the other person. Would your partner accept? If she were perfectly rational, she would. After all, it is a free hundred bucks. Surprisingly, though, in experiments such as this, people will turn down the offer and reject the money. In fact, they will reject it when the uneven split is 70/30 or more (Camerer, 2003). Why should this be? According to evolutionary theory, it is because people perceive the split to be unfair. Reciprocal altruism implies that if somebody does you a favor, then he or she is trustworthy. In the future, you could then return the favor and between the two of you establish a support network of mutual aid. This is one mechanism by which hunter–gatherer societies could establish strong bonds and increase their chances of survival.

We can effectively sum up this principle as "I'll scratch your back if you scratch mine" (Shermer, 2008). People are willing to accept help only if they perceive the split to be something approaching parity because this would signal an ally who could help them down the road. Interestingly, the effect isn't found only in humans. De Waal (1997b) had pairs of capuchin monkeys perform a task. When one member of the pair was rewarded and did not share with its partner, the other expressed displeasure and failed to cooperate from that point on. This and other evidence suggest that our primate ancestors and we have evolved a sense of justice.

The principle of **loss aversion** states that we are more sensitive to losses than to equivalent gains. In fact, empirical estimates show that people tend to fear losses about twice as much as they desire gains. Here's an example: Would you prefer to receive a $5 discount or to avoid a $5 surcharge? Most people prefer the latter. Here is another example: You are on a team of experts that must decide what course of action to take in preventing a flu outbreak.

> *Program A:* Two hundred people will be saved.
>
> *Program B:* There is a one-third probability that 600 people will be saved and a two-thirds probability that no one will be saved.

Of the participants, 72% chose Program A. Now consider the following rewording of the question:

> *Program C:* Four hundred people will die.
>
> *Program D:* There is a one-third probability that nobody will die and a two-thirds probability that 600 people will die.

The outcomes in this second case are identical to that in the first, yet now the responses are reversed. Of the participants, 78% now chose Program D over Program C.

These results demonstrate not just loss aversion but also **framing effects**. We prefer choices that are framed positively rather than negatively (De Martino, Kumaran, Seymour, & Dolan, 2006; Tversky & Kahneman, 1981). Framing effects are salient when we make financial decisions. Here is an example:

> *Query A:* The MePhone is available at Awesome Electronics for $100; five blocks away at DigiData, it is at half price for only $50. Would you make the short trip to save the $50?
>
> *Query B:* PowerPurchase offers the NetBook for $1,000; five blocks away, it can be had for $950 at CircuitCrazy. Would you make the short trip to save the $50?

The majority of people, when presented with Query A, answer "yes." They would be willing to walk the five blocks in that scenario. However, with Query B, even though the amount of money saved by the five-block walk is the same, many people respond "no."

Why? According to the principle of **mental accounting**, we put money into different categories based on the context or frame. We tend to put smaller amounts of money into one category and larger amounts into another. A $50 savings seems greater compared with $100 than it does relative to $1,000. Generally, people are less likely to make efforts to save cash when the relative amount of money they are dealing with is large.

Samuelson and Zeckhauser (1988) offered participants a choice between several financial investments varying in risk. Under those circumstances, individuals chose a variety of different investments based on how risk averse they were. Then, in a second condition, they told the participants that an investment had already been chosen for them, and they could either stick with this or switch to one of the alternatives. Now, they were far more likely to stick with the one that they had been assigned. Of these participants, 47% stayed with the default compared with the 32% of those who had chosen it in the earlier scenario. The tendency to stay with what we have (the status quo) and to prefer it over other options has been dubbed the **endowment effect**.

The evolutionary explanation for the endowment effect is straightforward. Back in the EEA, resources were usually scarce. Things such as food and tools took a long time and lots of effort to obtain. For this reason, it made sense to hoard or hold on to those things we already had, since replacing them would have been time-consuming and energy intensive. Those individuals who kept a tight grip on what they had would have been more likely to survive. However, the endowment effect can also backfire. Sometimes, it does make sense to let go of what we have and switch to something different that may end up being more valuable or at least less costly. Unfortunately, we tend to overestimate or focus on the amount of effort we've already put into something (its sunk cost) and use this to justify staying with it, which may blind us to choosing a better option. If this happens, then we have fallen victim to the **sunk-cost fallacy** (Sutton, 1991). Think for a minute about those things you may have held on to for too long. They may include staying with the same car, house, job, relationship, and so on.

Sex Differences in Cognition

Married couples may sometimes be aware of the existence of cognitive differences between the sexes. Perhaps one partner gets lost, while the other is forgetful. Research has shown that in the human species, there are some small but consistent cognitive differences between males and females. The evolutionary psychology approach tries to provide explanations for these differences in terms of the selection pressures that were acting on the two sexes during the EEA.

According to evolutionary psychology, important cognitive sex differences can be attributed to an early human **sexual division of labor**, according to which men hunted and women gathered (Silverman & Eals, 1992; Silverman & Phillips, 1998). The hunting process, which was apt to require navigation across long distances and novel terrain for days at a time, is believed to have selected for enhanced spatial abilities in men. Women, for whom communication with one another and with their children was a part of the gathering process, are believed to have developed superior verbal abilities. Indeed, such differences between the sexes are observable today. On average, males

outperform females on tests of spatial relations and image rotation (Levy & Heller, 1992). In these tests, a participant is usually asked to determine whether a letter of the alphabet, displayed in a variety of orientations, is the same as another. Females generally score higher than males on tests of verbal fluency, reading speed, reading comprehension, and spelling (Notman & Nadelson, 1991). One might also deduce that as women were specialized at gathering, they should be adept at specific kinds of spatial tasks—namely, object–location memory. Memory for object location is imperative to the task of gathering, as one would need to recall the locations of perhaps widely disseminated food sources. Silverman and Phillips (1998) discovered that women outperform men in their ability to remember the locations of objects, both natural and artifactual (see Figure 8.10). Their study suggests that our conceptions of spatial and verbal abilities are crude and that perhaps multiple domains within these abilities perhaps may point the way to additional sex differences.

A second theory that accounts for the differences between men and women with respect to spatial and verbal abilities has been advanced. According to this hypothesis, spatial ability became enhanced in males not as a result of hunting activity but, rather, as a result of long-standing mate-seeking practices. The males of polygynous animal species must travel comparatively greater distances to meet and mate with females. Data indicate that the males of polygynous species—for example, the meadow vole—make fewer errors than females when given maze navigation tasks (Gaulin & FitzGerald, 1986, 1989). These sex differences disappear in the prairie vole, which is a monogamous species and of which neither sex ranges very far. It should be noted that there is nothing special about "maleness" or spatial ability. Rather, it is the environmental demands that are placed on one sex or the other of a particular species that are pivotal. In the brown-headed cowbird, there is a greater need for spatial problem solving among females, who show evidence of spatial ability that is superior to that of the males (Sherry, Forbes, Khurgel, & Ivy, 1993).

That said, it is important to be careful with claims about the evolutionary origins of human cognitive abilities, in general, given the fact that there is usually very little concrete evidence to back them up. The evolutionary scholar Stephen Jay Gould (1978) called these claims "just-so stories" because they sound intuitively compelling but have little or no evidence, and an alternative invented story can often sound just as compelling. For example, Western thought has tended to assume that the reason men dominate in mathematical and scientific careers is because boys innately place more focus on objects and their mechanical and mathematical relationships, whereas girls innately focus on emotions and personal relationships (Baron-Cohen, 2003; Pinker 2003). On the face of it, this sounds like it might be a reasonable explanation. But the evidence actually does not support this just-so story. Developmental psychologist Elizabeth Spelke (2005) points out that if these sex differences are due to genetic biases, then we should be able to find evidence for them in infants. Although some studies show that older children and adults exhibit a sex difference in speeded quantitative reasoning tests, studies with infants do not show any advantage for males in processing objects, space, or number. Thus, many of the sex differences that show up in older children and adults may be due to cultural biases in how these skills are taught, rather than genetic biases.

Figure 8.10 A stimulus set that can be used to test for object and location memory. If you studied this for a minute, would you be able to recall what objects were next to the pelican?



EVALUATING EVOLUTIONARY PSYCHOLOGY

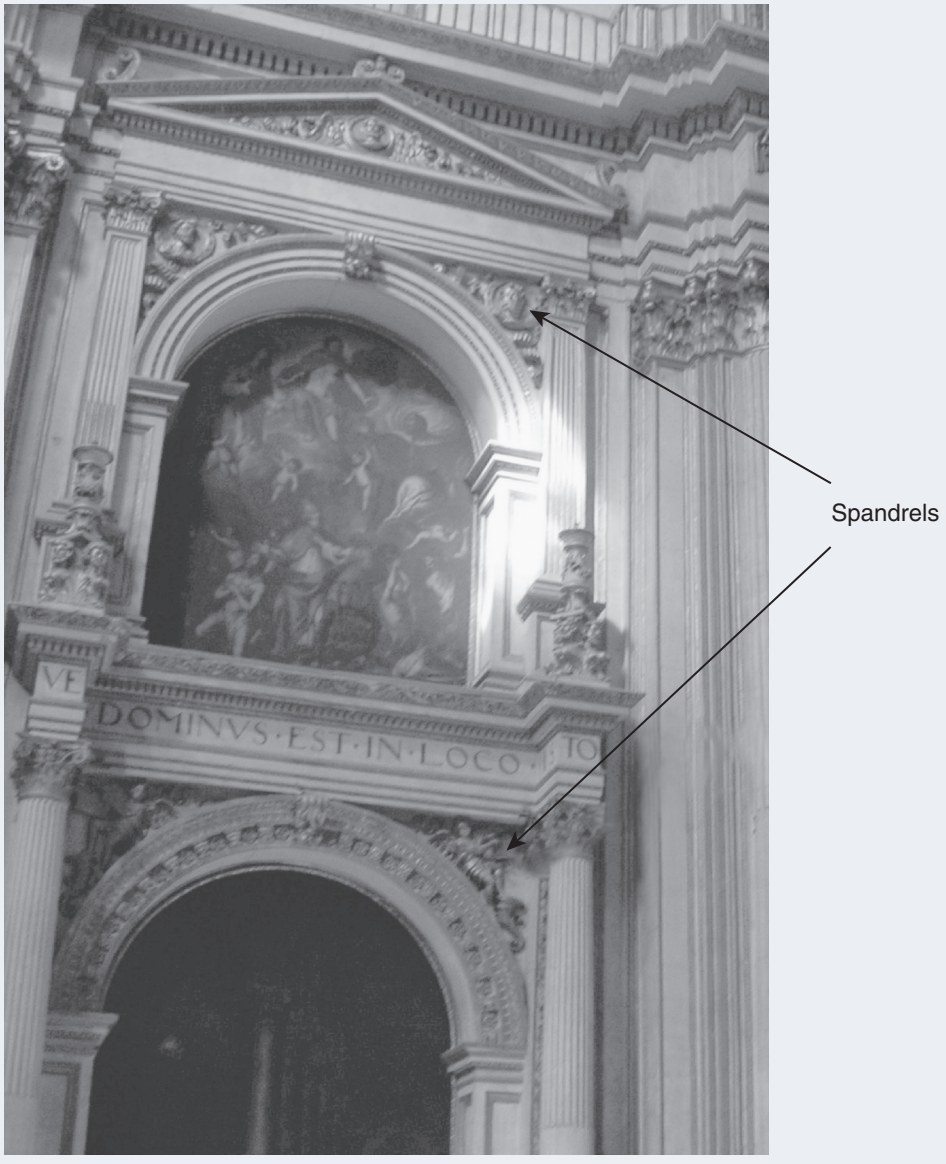
EP as a formal discipline is a recent addition to the ranks. As such, it has received both fervent acclaim and fervent criticism. Its strengths are its methodological variety and its attempt to develop an objective science of human behavior that is grounded on a single, powerful theoretical model. The evolutionary model already has proved to be fruitful in biology, and its use in psychology links the two disciplines. The evolutionary perspective also offers to reshape psychology such that it becomes a unified discipline, as opposed to the disjointed one that it tends to be currently (Buss, 1999). Modern psychology consists of multiple disciplines, each characterized by a unique theoretical perspective and a unique focus. For example, cognitive psychology focuses on information processing, social psychology focuses on interpersonal relationships, and developmental psychology focuses on growth and change in an organism over its entire life span. EP cuts across these different theoretical perspectives and tries to provide a coherent, unified approach to the study of mind and behavior. But EP has its problems. Even adherents of the theory of natural selection and of biological evolution have attacked its assumptions.

One of the fundamental assumptions behind the evolutionary approach is that evolved psychological mechanisms are adaptive responses to survival-related or reproduction-related problems. But this need not always be the case. At least two other kinds of phenomena, in addition to selective processes, can produce novel biological function (Dover, 2000). The first of these is **exaptation, or neutral drift.** In this instance, random mutations in an organism's genome result in new genes that have little or no consequence for reproduction. These genes can become disseminated in a population and at some point may assume a new adaptive function that represents a response to a change in the environment. The second is **molecular drive**, in which a copy of a gene can mutate. There are known cases in which genes that code for proteins become duplicated; the duplicated genes then diverge to serve another function. These genes and the characteristics they give rise to have not been selected for.

Stephen Jay Gould (2000) proposed that many of an organism's attributes may arise via nonselectionist mechanisms. He refers to these attributes as "spandrels." A **spandrel** is the triangular space between the exterior curve of an arch and the rectangular frame that encloses the arch (see Figure 8.11). A spandrel serves no design purpose; it exists solely as a consequence of an architect's surrounding an arch with a frame. Similarly, many psychological mechanisms may themselves be spandrels—they may be by-products of an adaptation that was subsequently co-opted for useful purposes. The ability to read and write are probably by-products of an earlier linguistic capacity—the capacity to comprehend and produce speech. Once these earlier capacities are in place, neural connections to visual and manual parts of the brain that were already in existence would then have enabled the later arriving capacities. Cognitive scientist Michael L. Anderson (2014) calls this "neural re-use," where perceptual and cognitive mechanisms that were naturally selected for in our ancestors now get used for new and different learned functions—often for functions that have not been around long enough for natural selection to have evolved genes for them in our DNA.

The job of the evolutionary psychologist is a difficult one, as he or she must start with a given mental mechanism and come up with a purpose for it. The purpose

Figure 8.11 Spandrels in a cathedral in Granada, Spain.



corresponds to the adaptive problem the mechanism was intended to solve. This process is called **reverse engineering** because one begins with the final product and thinks back to what it was that may have compelled its genesis. This is more difficult than the conventional engineer's job, which consists of starting with a desired function (getting a car

over a valley) and designing a product (a bridge) that will enable that function. It is also the case that evolutionary psychologists can easily be wrong in making the assumption that a particular psychological mechanism came into being as a response to a particular problem. It may have evolved in response to an entirely different problem.

What makes reverse engineering so difficult is the paucity of available evidence with which one might reconstruct crucial past events. Tools, bones, and other fragments of the archeological record allow us to make certain inferences, but they are insufficient to inform us about crucial aspects of prehistoric times—for example, kinship relations, group size, social structure, and the activities of males and females (Gould, 2000). These aspects of culture and behavior cannot be easily revealed through the fossil record.

In the traditional evolutionary view, biology is shaped by selection. Biology then gives rise to psychology. Psychology, in turn, generates culture—the larger environment in which humans live (Cosmides & Tooby, 1992). This scheme can be turned upside down, however, in that culture can also shape psychology. A large body of evidence shows that environments exert a strong influence on our thoughts and behaviors. Culture, primarily via the medium of learning, is thus another force that is capable of generating psychology.

Dynamical or **ecological models of cognition**, because they postulate the existence of learning, pose a threat to evolutionary psychology (Karmiloff-Smith, 2000). In these models, mental abilities arise through learning processes; that is, they come into being as a result of an individual's social, verbal, perceptual, and manipulative interaction with his or her environment—all taking place within the life span of the individual. In this account, what would seem to be an evolved psychological mechanism may, in fact, have been acquired by the individual as he or she went about his or her business. This runs contrary to the evolutionary view that natural selection pressures evolved psychological mechanisms that are coded for by genes that are present at birth, fully formed and just waiting to be activated by the appropriate stimulus inputs.

An alternative account has genes laying down the groundwork for a capacity that develops at a later stage as a result of experience. Stereoscopic vision, the ability to perceive depth via a combination of the information received from the two eyes, is an instance of this kind of capacity. Stereoscopic vision is not present at birth and develops fully only after an organism has perceived the world with both eyes. The brain mechanisms that underlie such abilities cannot exist in their entirety at birth, as developmental experience forges their basic neural connections and circuits—a process termed *postnatal neural assembly*.

Karmiloff-Smith (2000) distinguishes between two developmental models. In the **mosaic model of development**, a brain mechanism is determined almost entirely by genes and operates quickly (i.e., is a reflex or a fast perceptual process), and its components or parts develop independently of one another. Such mechanisms operate well under typical conditions—those "designed" to trigger their operation. They must, however, be completely specified in advance and are limited in how complex they can be. Evolutionary psychologists believe that evolved brain mechanisms are, in a word, mosaic. In contrast, in the **regulatory model of development**, brain mechanisms are only partially determined by genes and operate more slowly (i.e., they are deliberative cognitive processes), and their parts develop interdependently. This type of mechanism

is flexible and capable of altering its function under varying conditions, and only the broad outlines of such a mechanism need be genetically specified at birth. These mechanisms are also more complex. Research suggests that regulatory development is the norm for the growth of cortical abilities—those that underlie higher-order cognitive abilities. Mosaic development occurs in many animal species, and in humans, it is likely to govern several subcortical brain areas.

OVERALL EVALUATION OF THE EVOLUTIONARY APPROACH

Evolution is a powerful theoretical framework. It can be used to explain how virtually every natural process comes into existence. As such, it rests squarely within Marr's computational level of description, which attempts to provide explanation for a given process by identifying its purpose. Knowing why a mechanism exists gives us many clues as to its structure and function and can help us decipher problems that come into view at other descriptive levels. The evolutionary approach, as we noted earlier, meshes well with neuroscience. Collaboration between these disciplines should focus on finding the neural bases of evolved psychological mechanisms and on providing a fuller understanding of the selection pressures that created them.

One bone of contention between evolution and the cognitive and artificial intelligence perspectives is the issue of general-purpose versus domain-specific processing. Is the mind a general-purpose computer or a "Swiss army knife"? As is the case with many other debates that are apt to pit the proponents of one extreme against those of the opposite extreme, it may be that both are correct. Some cognitive processes may indeed qualify as evolved psychological mechanisms, by their being in accordance with all the defining criteria. Others may be of a general purpose. There is a need for studies that will focus on this important distinction and on the role of complex developmental factors.

SUMMING UP: A REVIEW OF CHAPTER 8

1. Biological evolution is based on the theory of natural selection that describes how species change over time. Three principles are important: (1) variation in a species population, (2) inheritance of traits, and (3) a change in the environment favoring some traits over others, known as natural selection.
2. The field of comparative cognition examines the thinking capacities of animals and attempts to make cross-species comparisons. Specific animal species adapt to their ecological niches and, depending on the circumstances, can evolve similar or differing perceptual, conceptual, memory, and reasoning abilities. However, there are numerous problems in making these comparisons due to different methodologies and anthropomorphism.
3. EP is a field that examines the role that evolutionary forces played in determining our cognitive abilities. The EEA is the period

during which such forces played a dominant role.

4. The classical view sees cognitive processes as general purpose and able to solve a variety of different problems. The evolutionary psychology view instead sees cognitive processes as special purpose, having evolved to solve a particular adaptive problem. Such processes are called evolved psychological mechanisms.

5. There is evidence to suggest that evolutionary forces have shaped all cognitive processes, including categorization, memory, logical reasoning, judgment under uncertainty, and language.

6. Humans rarely solve problems in a rational or objective fashion. Rather, they rely on heuristics. These are strategies that provide a fast and easy solution but are not always correct. Heuristics can lead them to commit fallacies, or fundamental misunderstandings, of the logical or mathematical rules underlying a problem.

7. Examples of heuristics and fallacies abound. These include the representativeness heuristic, the base-rate fallacy, the conjunction fallacy, and the gambler's fallacy.

8. Behavioral economics is a relatively new discipline that examines how we reason about financial matters. It turns out that in this domain, we also use heuristics and fall victim to fallacies. In the ultimatum game, we reject offers of money if they are judged to be unfair. We are loss aversive, are sensitive to the way questions are framed, think differently about large and small sums of money, and tend to hold on to what we have even if it might end up being more costly than other alternatives.

9. Evolutionary pressures can help explain differences in the way males and females think. Male hunting and female gathering could account for male superiority in certain spatial tasks and female superiority in particular verbal tasks.

10. EP has many critics. Processes other than natural or sexual selection, such as exaptation (or neutral drift) and molecular drive, can account for the development of human mental faculties. It is also difficult to determine what the conditions were like during the EEA since we cannot observe it directly. Even so, the evolutionary approach has much promise, as it may serve to unify disparate fields in psychology and cognitive science.

SUGGESTED READINGS

Adami, C. (1999). *Introduction to artificial life*. New York, NY: Springer.

Buss, D. M. (2007). *Evolutionary psychology: The new science of the mind*. Boston, MA: Allyn & Bacon.

Dawkins, R. (2006). *The selfish gene*. Oxford, England: Oxford University Press.

Halpern, D. F. (2000). *Sex differences in cognitive abilities*. Mahwah, NJ: Erlbaum.

Kahneman, D., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty*. Cambridge, England: Cambridge University Press.

Rose, H., & Rose, S. (Eds.). (2000). *Arguments against evolutionary psychology*. New York, NY: Harmony Books.

CHAPTER NINE

THE LINGUISTIC APPROACH

Language and Cognitive Science

THE LINGUISTIC APPROACH: THE IMPORTANCE OF LANGUAGE

Linguistics is the study of language. It played a key role in the initial formation of cognitive science in the 1960s, especially with Noam Chomsky's (1959) critique of the behaviorist account of language learning in B. F. Skinner's 1957 book, *Verbal Behavior*. Rather than accept a behaviorist account that implies you can only learn to say things that you've heard other people say, Chomsky pointed out that many of the sentences we produce are sentences that we've never heard before. Linguists helped frame cognitive science at that early point in time as an interdisciplinary field that was *not* rooted in psychology but instead combined equally the efforts of computer science, linguistics, philosophy, psychology, and neuroscience, as well as some anthropology, economics, and education. There are many different kinds of linguistic frameworks—each with its own theoretical perspectives and methodologies. Some of these adopt a neuroscience approach and use neuroimaging methods or study the language-related deficits of patients who have suffered brain damage. Others implement various network models of how language information is represented and processed. Some linguists take on a developmental orientation: They examine how language ability grows and changes with time during the development of the individual. Still others who study linguistics are philosophers who ask questions about the nature of language and the relationship between language and thought. In fact, language can be studied from the vantage point of every field of study that has been described in this book and more. What makes linguistics unique, then, is not the perspective or the tools it brings to the table but the subject matter of the investigation—language itself.

Linguistic studies attempt to answer many questions. Language is so complex that much of the research that has been conducted

Learning Objectives

After reading this chapter, you will be able to:

1. Distinguish phonology, morphology, syntax, and semantics.
2. Understand how to interpret a speech spectrogram.
3. Describe what a syntactic ambiguity is in sentence processing.
4. Describe what an image schema is in cognitive linguistics.
5. Explain the negative evidence problem in language learning.
6. Contrast domain-general and domain-specific mechanisms for language acquisition.
7. Define the linguistic relativity hypothesis.
8. Identify brain regions associated with language comprehension and production.

in this area has been directed toward an understanding of the structure of language (in addition to how it is used). These studies have focused on grammatical rules that specify allowable combinations of linguistic elements. Another interesting issue is whether humans are unique in using language or whether some animals possess language ability. Languages, of course, cannot be learned overnight, and many linguists have studied language acquisition—how it is acquired during development. Linguistics is truly interdisciplinary, and in this chapter, we will survey the approaches to the study of language that are represented by the disciplines of linguistics, philosophy, cognitive psychology, neuroscience, and artificial intelligence.

THE NATURE OF LANGUAGE

With this in mind, we begin our discussion of the linguistic approach with an exploration of the nature of language. There has been much debate about what language is, exactly—and there is no agreed-on definition. It is easier to list its most important characteristics (in lieu of providing a definition). According to Clark and Clark (1977), language has five characteristics:

1. *Communicative:* Language allows for communication between individuals. **Communication** refers to the production, transmission, and comprehension of information.
2. *Arbitrary:* A language consists of a set of symbolic elements. Symbols, as we noted in Chapter 1, are referential—they stand for or refer to something. Linguistic symbols can be almost anything. Most commonly, these symbols are for sounds, pictures, or words. The defining hallmark of these symbols is that they are **arbitrary.** Virtually any sound, picture, or word could be chosen to represent a particular thing. For example, native speakers of American English may feel as though the spoken word *truck* fits well in its reference to a large heavy mechanical 18-wheeler. Then, they may be surprised when they first learn that, in England, that same imposing metallic vehicle is called a "lorry." It almost sounds like the wrong kind of name for such an object. Linguistic symbols indeed tend to be arbitrary in how they make reference to objects and events. (However, there are some occasional exceptions to this arbitrariness [Bergen, 2004; Farmer, Christiansen, & Monaghan, 2006]. For instance, in English, the speech sound "gl" tends to be at the beginning of a lot of words that refer to light and vision, such as glimmer, glisten, glitter, gleam, and glow. These **phonaesthemes** show that language is not completely arbitrary all the time.)
3. *Structured:* The ordering of the symbols in a language is not arbitrary but is governed by a set of rules, or is **structured.** The rules specify how the symbols may be combined. In English, we place the adjective before the noun, as in the phrase "the big house." In Spanish, this same phrase is expressed via a different set of rules: The adjective follows the noun, as in "la casa grande."

4. *Generative:* The symbolic elements of a language can be structurally combined to create a nearly infinite number of sentences. Just think of how many six-word sentences one would be able to generate in English. If we start with the sample sentence "The fox jumped over the fence," we can then substitute the words *dog*, *cat*, *deer*, and words for many other animals in place of *fox*. Likewise, we can substitute the words *bottle*, *can*, or *tire* for the word *fence*. So the number of variations on just this one sentence is large. Every day, we utter new sentences that we never have uttered before. The **generative** property of language makes language very powerful, as almost any idea that can spring to mind can be expressed.
5. *Dynamic:* Languages are **dynamic**, constantly changing as new words are added and grammatical rules altered. Only 20 years ago, the word *tweet* referred merely to the chirp of a small bird—because Twitter didn't exist.

There is a fundamental distinction to be made that has to do with the type of linguistic representation: whether it is auditory, visual, or having to do with another sensory domain. Spoken language is naturally produced via the faculty of speech and is understood via listening. Speech and listening to speech can, of course, be transformed into their approximate equivalents within the visual domain—writing and reading. However, a number of "extralinguistic" information sources are lost when natural conversation is converted into written text. Speech intonation (e.g., prosody) can allow the raising and lowering of vocal pitch to change the meaning of a sentence (Brown, Salverda, Dilley, & Tanenhaus, 2011). Moreover, manual gestures (such as pointing, iconic shapes, and rhythmic movements) also add meaning to spoken sentences. Beyond this, there are also languages that consist entirely of manual gestures, such as the American Sign Language (ASL), and tactile languages, such as Braille. We focus here on spoken and written language as they are the most widely used forms of language processing.

LANGUAGE PROCESSING

The study of language processing is roughly carved up into short, medium, and long time scales. Happening at the shortest time scale, over the course of a few dozen milliseconds, is phonology. **Phonology** is the study of how we produce and perceive individual speech sounds (phonemes), such as the "v" sound in the word *vote*. At a slightly longer time scale of hundreds of milliseconds, **morphology** studies how those meaningless phonemes are combined to make words that have meaning. At the longer time scale of seconds, syntax is the study of the *structure* (or grammar) of phrases and sentences, and **semantics** is the study of the *meaning* of those phrases and sentences. At an even longer time scale, **pragmatics** is the study of the social purpose and intent of those sentences.

Phonology

A **phoneme** is the smallest unit of sound in the sound system of a language. In general, a phoneme has no meaning on its own (except for the occasional phonaestheme).

Phonemes correspond in a very rough way to the letters of an alphabet; in some instances, multiple phonemes correspond to a single letter. The phoneme for the letter *a* as it is pronounced in *father* corresponds to the "ah" sound, whereas a different phoneme for the letter *a* as it is pronounced in the word *cane* corresponds to the sound "ay." There are about 44 phonemes in the English language. In Table 9.1, 59 consonant phonemes are shown from the International Phonetic Alphabet. When vowels are included, the number of phonemes used across languages worldwide is over 100, but most languages only use 30 to 50 of them. In the top-left corner of Table 9.1 are the phonemes /p/ and /b/, as in the words *pear* and *bear*, respectively. The place of articulation is at the touching of the two lips (bilabial), and the manner of articulation is called plosive because the lips stop the air completely and then release it in a burst. The /p/ phoneme is called unvoiced because the vocal cords are not vibrating during the production of this phoneme, whereas the /b/ phoneme is called voiced because they are. Touch your Adam's apple while saying "bear" and "pear" very slowly, and you will notice that your vocal cords are vibrating during the /b/ in "bear" but not during the /p/ in "pear." Halfway down the second column of Table 9.1, the same thing is seen with the unvoiced /f/ phoneme and the voiced /v/ phoneme. Those two phonemes happen at the same place of articulation, with the lip pressed against the teeth (labiodental) and with the same manner of articulation as well, straining air out of the mouth through narrowed openings (fricative). The next columns describe different places in the mouth where the tongue touches to produce the articulation.

Table 9.1 Consonant phonemes from languages around the world.

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2015)

CONSONANTS (PULMONIC) © 2015 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k g	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

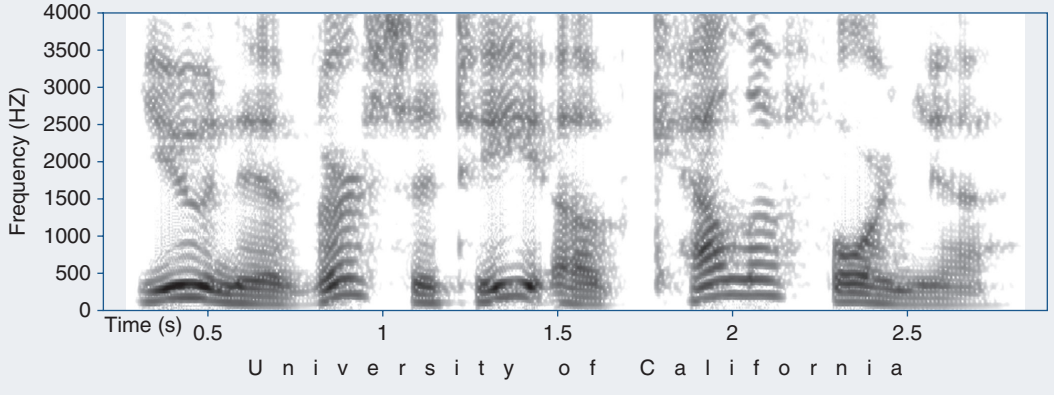
Source: IPA Chart, http://www.internationalphoneticassociation.org/content/ipa-chart, available under a Creative Commons Attribution-Sharealike 3.0 Unported License. Copyright © 2015 International Phonetic Association.

A **speech spectrogram** is a visual representation of the speech signal; it is a graph that displays the amount of energy in low, medium, and high frequencies of a speech sound over time (see Figure 9.1). Note how the vowel phonemes in Figure 9.1 have resonant bands of energy at the lower frequencies (the curvy dark stripes). The consonant phonemes have less resonance. Note how the fricatives "s" and "f" in "University" and "California" both have energy only in the higher frequencies, producing mostly a hissing sound. Now compare the "v" in "University" with the "f" in "California" to see the resonant low-frequency voicing in the /v/ phoneme and the absence of voicing in the /f/ phoneme (just as described in Table 9.1).

In speech perception, mapping phonemes onto words is difficult for two main reasons. The first of these concerns word boundaries. It turns out that there are often no pauses between words in spoken speech. In Figure 9.1, note how there is no silence between "University" and "of." The "y" in "university" blends into the "o" in "of," making one smoothly curved vowel formant. Moreover, there is a silence internal to the word *University*. Immediately after the final "i" in "Universi" and before the "t" in "ty," there is a brief silence as the tongue closes the flow of air to make the plosive /t/ phoneme. Therefore, if one tried to segment speech into words by looking for the silences, one would make a lot of mistakes (Saffran, Aslin, & Newport, 1996). The second major issue is phoneme variability. If each phoneme were pronounced clearly and uniformly, speech recognition would be much easier. This is not the case. Speakers vary tremendously with respect to the pitches and durations of their phonemes. They are apt to pronounce a given phoneme variably and in ways that depend on which phonemes precede it and which come after it. This is known as **coarticulation**. Additional complicating factors are the presence of background noise and the fact that in English, a single speech sound—for example, *bear* and *bare*—can be ambiguous, belonging to more than one word.

People resolve these ambiguities by taking into account the surrounding context of the sentence. In one classic study, Richard Warren (1970) presented participants with

Figure 9.1 A speech spectrogram.



recordings of sentences in which a phoneme was removed and replaced with a brief cough sound. One of the sentences was "The state governors met with their respective legi*latures convening in the capital city." (The asterisk indicates where the cough sound was inserted.) Essentially, all participants reported that the sentence sounded fully intact, with no missing speech sounds, and when asked to mark on a written version of the sentence where the cough took place, accuracy was poor. It was as if the missing phoneme was "filled in" by the listener's brain. Further examples of context assisting speech perception come from replacing a phoneme in a sentence with a synthetically blended version of two phonemes. Take a spoken sentence like, "The moon rises at t/d-usk," where the t/d sound is the synthesized blend of the voiced /d/ phoneme and the unvoiced /t/ phoneme (near the top middle of Table 9.1). Listeners routinely experience the illusion of hearing the word as "dusk" in that sentence, with no uncertainty or confusion. However, when the same ambiguous blended word is presented at the end of "The walrus was missing a . . .," listeners then clearly and confidently perceive that word as "tusk" instead (Connine, 1987). In fact, neuroimaging research has shown that, when this speech perception illusion happens, left temporal brain regions associated with spoken word recognition are influencing left frontal brain regions associated with phoneme perception to make it happen (Gow & Olson, 2016; see also Getz & Toscano, 2019; McClelland, Mirman, & Holt, 2006). This clearly demonstrates that the meanings of words in a sentence that have already been understood provide a framework for understanding the words that are just now being heard. It also shows that in human speech perception, recognition is top-down as well as bottom-up, as the meaning of the entire sentence is pulled together simultaneously with the meanings of individual words.

Humans also have the benefit of visual cues when they are taking in speech. We can look at a speaker's lips as he or she is speaking. The positioning of the lips can influence how we interpret phonemes. In 1976, McGurk and MacDonald voice-dubbed the repeated syllable "ba-ba-ba-ba" over a video of someone actually saying "ga-ga-ga-ga." When listeners watch this video, they often feel like they are hearing "da-da-da-da" because the lips are not touching in the video (which would be required for an actual "ba-ba-ba-ba" sound). This audiovisual speech perception illusion is influenced both by synthetically blended phonemes and by synthetically blended lip shapes on an animated face (Massaro, 1998). Moreover, neuroimaging findings suggest that when the visual context successfully alters the perception of the phoneme, this effect is happening at the earliest stage of speech perception in auditory cortex (Shahin, Backer, Rosenblum, & Kerlin, 2018).

Morphology

As those phonemes get perceived one after another (with all manner of contextual help), they must be strung together correctly to recognize words. **Morphemes** are structured collections of phonemes that make up the smallest units of spoken language that have meaning. Morphemes very roughly correspond to words but can also be the parts of words. Thus, the sound of the spoken word *apple* is a stem morpheme, but the sound of *-s* denoting the plural form is a bound morpheme because it can only be used when it is bound (or attached) to a stem morpheme. If we want to change the form of *apple*

from singular to plural, we add the bound morpheme *-s* as a suffix to form *apples*, which changes the meaning. Similarly, there is the bound morpheme that corresponds to the sound *-ed*, which, when added as a suffix to the root form of many verbs, forms the past tense; for example, changing *vote* to *voted*. Another example bound morpheme is *-ment*, which can change a verb to a noun when added as a suffix, thus changing *establish* to *establishment*. An additional bound morpheme can be attached to this word, such as *-arian*, to make the word *establishmentarian*—which refers to a person who supports the established church. English also allows bound morphemes to be prefixes, attaching to the front of a stem morpheme. Thus, the bound morpheme *dis-* can be added to the front of that word to make *disestablishmentarian*—which refers to a person who opposes the established church. Other languages, such as Latin, Sanskrit, and Arabic, have *infix* bound morphemes that can change the meaning of a word by being added into the *middle* of the stem morpheme. Considering that there are over 100,000 words in the English language, the number of morphemes that the language has is quite large. However, most English speakers only know about 20,000 to 40,000 words, and casual everyday conversation and writing tends to use only about 5,000 of those.

Word Recognition

The recognition of spoken morphemes, spoken word recognition, is a cognitive process that has been studied in great detail. In reaction time experiments, William Marslen-Wilson (1987) showed that, partway through hearing a word (such as *elephant*), mental representations of other words that start with the same first few phonemes (such as *elevator* and *elegant*) are primed or partially activated, as indicated by faster reaction times to those similar words. Essentially, those first few phonemes are temporarily ambiguous as to what word they will turn out to form. Thus, during the first couple hundred milliseconds of hearing a spoken word, those first few phonemes trigger the partial activation of several lexical representations (i.e., word meanings in the mind). Then, as the next few phonemes are heard, that collection of partially activated lexical representations winnows down to eventually just the word that was actually spoken (McClelland & Elman, 1986). For example, in eye-tracking experiments and computer mouse tracking experiments, when people are instructed to "click the candy," they often briefly look at and partly reach toward a picture of a *candle* before finally resting on the picture of the candy (Allopenna, Magnuson, & Tanenhaus; 1998; Spivey, Grosjean, & Knoblich, 2005). Thus, the spoken word recognition process does not wait until it has collected all the phonemes of a word, and fully recognized exactly the word it is hearing, before it sends that information out to other parts of the brain. Instead, well before the word is finished being heard, estimates of the multiple possible words that this current word could turn out to be are continuously shared with other cognitive and linguistic processes.

In spoken word recognition, one must hear one phoneme after another in sequence. However, in written word recognition, processing is somewhat different because the eyes do not fixate one letter at a time from left to right while reading. Rather, the eyes jump from word to word, often stopping only once on short words and perhaps twice on longer words (Rayner, 1998). During that steady eye fixation near the center of a short word, its letters are processed more or less in parallel (all together at the same time).

Words that have irregular pronunciations, such as *pint*, often take longer to process for reading aloud (Glushko, 1979). *Pint* is irregular because its pronunciation requires the long vowel *i*, whereas the majority of English words with those last three letters are pronounced with a short vowel *i* (e.g., mint, hint, and lint). Thus, the mapping of a set of letters (orthography) onto a set speech sounds (phonology) is a complex process that is influenced by similar-looking and similar-sounding words (Seidenberg & McClelland, 1989). However, even in silent reading, words that are difficult to process because they are infrequent (e.g., *convex*), or because they are ambiguous between two meanings (e.g., *bank*), will tend to be fixated for a longer time than common unambiguous words (Rayner & Duffy, 1986).

When we read or hear an ambiguous word, such as *bank*, which could refer to a financial institution or a riverbank (among other meanings), we quickly use context to determine which meaning is appropriate. But it is not necessarily instantaneous. It may take a few hundred milliseconds after hearing an ambiguous word for the listener to winnow down the possible meanings to only the contextually appropriate one (Tanenhaus, Leiman, & Seidenberg, 1979). During those few hundred milliseconds, the contextually irrelevant meaning may be partially active. However, with a context that strongly constrains the sentence toward the more frequent meaning of the ambiguous word, such "After work on Friday, I deposited my paycheck at the bank," some studies indicate that the contextually irrelevant meaning may not get significantly active at all (Kawamoto, 1993; Tabossi, 1988).

These examinations of the process of resolving an ambiguity have been instrumental in developing an understanding of the architecture of the language processing system. When early studies suggested that context may not be able to have an immediate influence on resolving an ambiguous word, theorists interpreted this as evidence for an encapsulated module for word recognition that maps letters and phonemes onto lexical representations but is unable to be influenced by the context from other information in the sentence (Swinney, 1979). However, once evidence was found that strongly constraining contexts can sometimes have an immediate influence, theorists proposed nonmodular architectures where subsystems for phonological, lexical, syntactic, and semantic processes might all be interconnected (Trueswell & Tanenhaus, 2005). This use of ambiguity resolution as a tool for exploring the cognitive architecture is also seen in the study of syntax.

Syntax

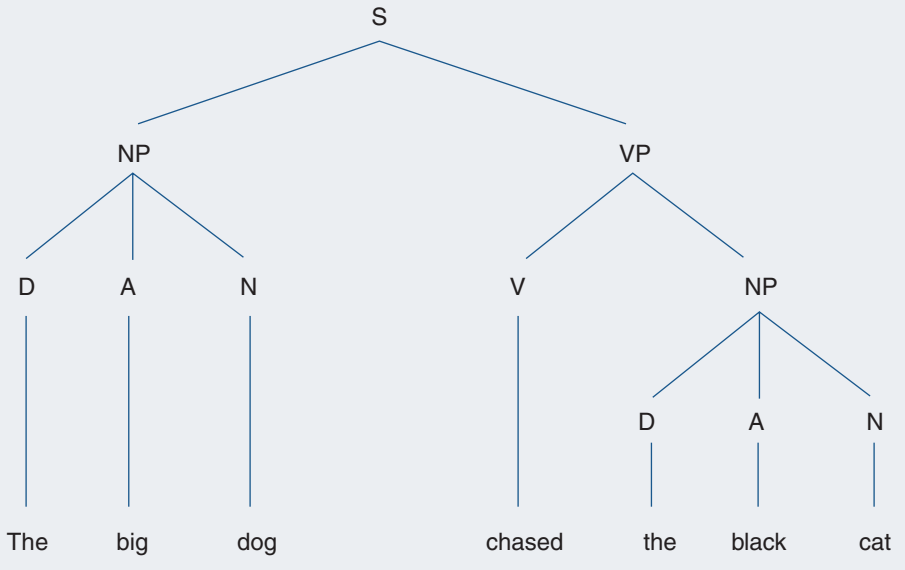
Linguist Noam Chomsky's most important contribution to linguistics was to bring attention to the complexity of syntactic **structure** in human languages. While there is some hierarchical structure involved in music, dance, and other behaviors, the sheer complexity of the syntactic structure in one lengthy sentence far outstrips the others. Syntax is important because it tells us what is a proper way of expressing something in a language and what is not. If there were no rules or constraints on expression, we could string words together in practically any order, and it would be impossible to convey anything.

Sentences have distinct parts that are arranged in a structural hierarchy. This organization is called a **phrase structure** and can be illustrated via the use of syntactic tree diagrams. Figure 9.2 is a syntactic tree for the sentence "The big dog chased the black cat." At the highest level of the hierarchy, the entire sentence (S) is represented. Moving down one level, the sentence is composed of two parts, a noun phrase (NP) and a verb phrase (VP). Moving down another level in the hierarchical structure, we see that the noun phrase is made up of a determiner (D), an adjective (A), and a noun (N). Determiners are words such as *a* or *the*. The verb phrase is made up of a verb (V) and another noun phrase that itself contains another determiner, adjective, and noun.

There is a grammar that governs the use of phrase structures. This **phrase structure grammar** imposes certain limitations on how a legitimate sentence can be put together. One phrase structure rule is that all sentences are composed of a noun phrase and a verb phrase. A second rule is that noun phrases consist of a determiner followed by a noun (with optional adjectives). Verb phrases can be expressed a bit more flexibly—as a verb followed by a noun phrase, a prepositional phrase, another sentence, or other elements.

As you read or hear a sentence, one word at a time, temporary syntactic ambiguities frequently show up—a little bit like to those found with word recognition. And just like with spoken word recognition, the language system does not wait until all the words have been delivered before it tries to figure out what to do with them. Upon reading

Figure 9.2 The syntactic tree for a simple sentence.



or hearing the first several words of a sentence, the language system is already trying to build a syntactic structure that might work for the sentences. In fact, that is precisely why temporary ambiguities arise. Figure 9.3 shows a partial syntactic tree for the first four words of the sentence, "Put the apple on the towel in the box." Upon encountering the preposition "on," the syntactic processing system is faced with an ambiguity: Its rules allow it to attach the upcoming prepositional phrase as part of the VP or as part of the NP. If the prepositional phrase is attached to the VP, then "on the towel" will be understood as the destination of the "put" action. As a result, the next prepositional phrase "in the box" may be hard to make sense of. It will feel as though the reader or listener has been "led down the garden path" to a dead end, and the sentence doesn't seem grammatically correct anymore. Sentences like these are called "garden-path" sentences (Frazier & Rayner, 1982).

To explore how these syntactic ambiguities get resolved, psycholinguist Michael Tanenhaus presented spoken versions of these sentences to people while tracking their eye movements to real objects on a table (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). He found that when there was only one apple on the table (as in Figure 9.4), people often revealed a brief "garden-path" effect by looking first at the apple (partway through hearing the sentence) and then *looking at the other towel* in the top right, before then looking at the box and finally placing the apple in it. Essentially, upon hearing "the apple," they knew which apple was being referred to because there was only one there. Thus, when the prepositional phrase "on the towel" came in, they assumed it was part of the "put" event (attached to the VP) and then looked at the other towel as a potential destination. That other towel in the top right is actually irrelevant to the meaning of the sentence, but for a short moment, many listeners understood it to be the destination for the "put" event. In a fraction of a second, they corrected that "garden path" and carried out the correct action.

This scenario contrasts with the alternative visual context that Tanenhaus et al. used. Instead of a single referent for the word *apple*, Figure 9.5 shows the two-referent visual context (containing two apples). In this context, when people heard "the apple," they

Figure 9.3 The syntactic tree for the initial portion of a syntactically ambiguous sentence.

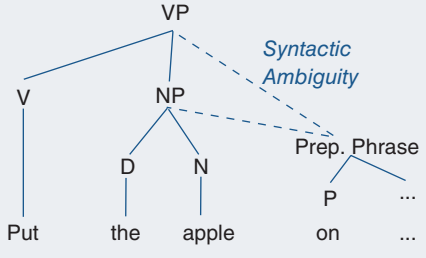


Figure 9.4 One-referent context for "Put the apple on the towel in the box" (Tanenhaus et al., 1995).

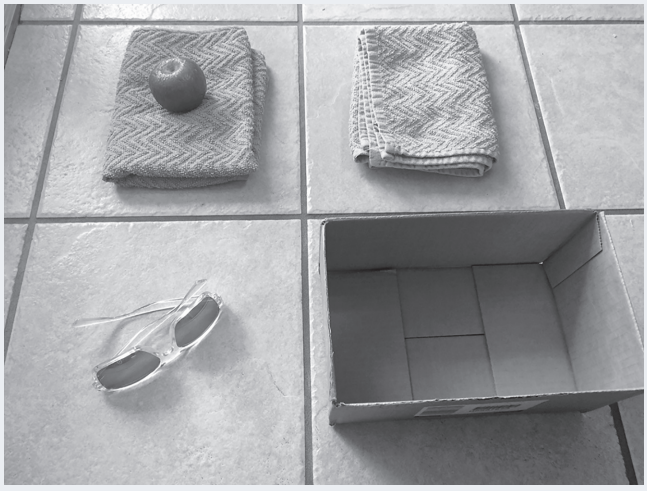
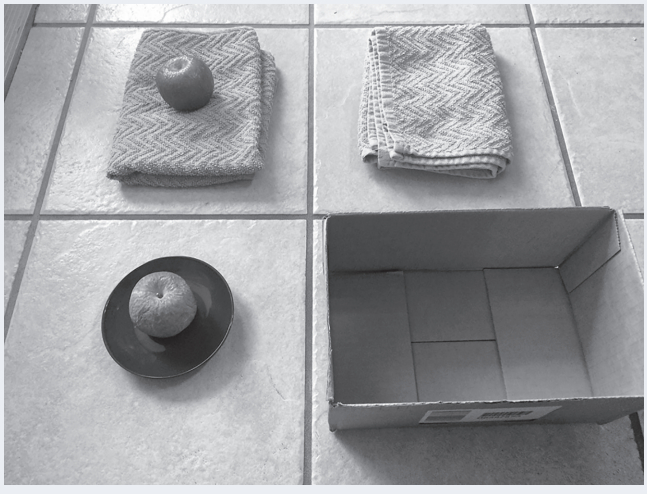


Figure 9.5 Two-referent context for "Put the apple on the towel in the box" (Tanenhaus et al., 1995).



did not immediately know which apple was being referred to; further specification of that NP is sorely needed. Thus, upon hearing the syntactically ambiguous prepositional phrase, "on the towel," the referential context made it highly preferable to attach it to the NP, providing that further specification (Altmann & Steedman, 1988). In this two-referent context, participants rarely looked at the irrelevant towel. Thus, visual context influenced syntactic processing so that listeners did not get garden-pathed by the sentence when heard in this context.

As was seen with word recognition, the use of ambiguity resolution as a tool revealed that syntax is not handled by an encapsulated module that only processes syntactic information. In addition to visual context, syntactic processing is also quickly influenced by statistical patterns in lexical representations (MacDonald, Pearlmutter, & Seidenberg, 1994), semantic information (McRae, Spivey-Knowlton, & Tanenhaus, 1998), phonological information (Farmer et al., 2006), and even constraints on potential actions in the display (Chambers, Tanenhaus, & Magnuson, 2004).

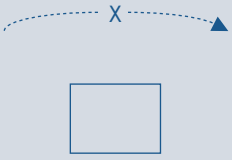
Semantics

Semantics focuses more on the *meanings* of words and sentences and less on how they are *structured*. Whereas Noam Chomsky placed syntax at the heart of language, George Lakoff placed semantics at that heart. In the 1970s, Lakoff and his colleagues developed a subfield of linguistics called cognitive linguistics that emphasizes how language is not processed inside a module that is encapsulated from the rest of cognition but instead processed in concert with other cognitive mechanisms—especially imagery and perception.

With the development of image schemas, cognitive linguistics pointed to visual imagery, visual perception, and spatial reasoning as fundamental mechanisms underlying linguistic meaning (Johnson, 1987; Lakoff, 1987). Image schemas are two-dimensional schematic diagrams that embody the meaning of various relational linguistic terms, especially verbs and prepositions. Thus, understanding the meaning of a preposition such as "over" may have underlying it an image schematic process that represents its meaning. Figure 9.6 shows an image schema for "over," where the X would be the subject of the sentence and the square would be the object (and the dotted line shows optional motion for certain verbs). According to image schema theory, an implicit mental representation of an image such as that is part of how one understands a sentence like "The pigeon flew over the house" or even "Mary went over her notes" (Brugman & Lakoff, 1988). Mary could very well be lying on her couch and holding her notebook above her. But conceptually and metaphorically, s*he* is still above *them*, "going over" them. Reaction time experiments indicate that the directions of these image schemas (for verbs like *run*, *sink*, and *increase*) have influences on visual perception and visual memory (Richardson, Spivey, Barsalou, & McRae, 2003). These image schemas may even play a role in early prelinguistic cognitive development (Mandler, 1992).

Cognitive linguist Leonard Talmy (2000) developed images schemas with movement and force dynamics to help define not only literal meanings but also metaphorical meanings. For example, in the sentence, "The road runs through the valley," nothing is actually running. Talmy called this figurative use of action verbs fictive motion. Although the sentence,

Figure 9.6 A sample image schema for the preposition "over."



"The road is in the valley," pretty much says the same thing, it doesn't use fictive motion. Interestingly, the small distinction between these two sentences, with one of them using fictive motion and the other not, causes people to understand the two sentences slightly differently (Matlock, 2004). In fact, when people look at a drawing of a road in a valley, their eye movements reveal that they are treating the road as if it is carrying out some kind of movement when they hear the fictive motion sentence (Richardson & Matlock, 2007).

An image schema can also embody abstract concepts inherent in its shape and layout, such as containment, support, and agency. Gibbs and O'Brien (1990) investigated people's mental imagery for various idioms, like "spill the beans," and found that idioms have highly consistent imagery across English speakers, whereas comparable nonidiomatic phrases, like "spill the peas," do not. They suggested that this consistency in mental imagery for the meanings of common idioms is due to image schemas giving rise to conceptual metaphors. When we say things like "His mind is *full* of ideas," or "My mind is *empty*," we are relying on an underlying metaphorical concept in which the mind is treated as though it were a container: A conceptual metaphor of MIND IS A CONTAINER (Lakoff & Johnson, 1980). Cognitive linguists propose that a great many conceptual metaphors (such as LIFE IS A JOURNEY and ANGER IS HOT FLUID UNDER PRESSURE) are at the root of how we reason about the world and how we use language. The reliance on the body as a source for many conceptual metaphors in cognitive linguistics helped pave the way for the field of embodied cognition (see Chapter 11).

Recall from the beginning of this chapter that semantics can "sneak into" phonology with the occasional phonaestheme, where a speech sound such as "gl" tends to be at the beginning of words that refer to light and vision, such as glimmer, glisten, glitter, gleam, and glow. As it turns out, semantics can also sneak into the grammatical gender system for languages that use male and female determiners (e.g., "la" and "el" in Spanish). Cognitive scientist Lera Boroditsky showed that linguistic gender is not as arbitrary as we often think—a little bit like phonaesthemes. She asked a group of native Spanish speakers (who were also fluent in English) and a group of native German speakers (who were also fluent in English) to write down in English the first three adjectives that came to mind for a set of concepts that were all presented in English (Boroditsky, Schmidt, & Phillips, 2003). For half of the concepts, the Spanish translation used feminine grammatical gender and the German translation used masculine grammatical gender. The reverse was true for the other half. Although the grammatical gender of nouns is generally considered an arbitrary grammatical

assignment that says nothing essential about the meaning of the noun itself, Boroditsky found that it influenced people's mental representations of those concepts. Take, for instance, the Spanish word *el puente* (i.e., *bridge* in English). The Spanish determiner "el" shows that the grammatical gender for this word is masculine, but in German, the very same object uses feminine gender, *die Brücke*. Boroditsky's findings revealed that grammatical gender wasn't entirely arbitrary after all because the Spanish speakers and German speakers wrote down features of the concepts that were significantly influenced by the gender marking. For instance, Spanish bilinguals described *bridge* as "big," "strong," and "towering," clearly placing emphasis on masculine-like qualities of bridges. By contrast, German bilinguals tended to describe *bridge* as "beautiful," "fragile," and "elegant." The reverse happened with the concept *key*, which is feminine in Spanish and masculine in German. Spanish bilinguals listed English adjectives such as "lovely," "little," and "intricate" for *key*, while German bilinguals listed English adjectives such as "jagged," "heavy," and "hard."

While semantics deals with literal and figurative meanings of phrases and sentences, there are also larger-scale goals of everyday language use that rest on top of that. When we carry on a conversation with someone, we are not solely trying to transmit meanings back and forth. Sometimes we have social intentions that involve convincing the person we're talking to (i.e., our interlocutor) to *do* something for us. Pragmatics is the branch of linguistics that addresses those social intentions of language use.

Pragmatics

Human language is a social construct. Individuals use it to communicate with one another. We communicate for all sorts of reasons. Sometimes, the purpose of a linguistic utterance is simple conveyance of information, as when we say, "This pillow is soft." But linguistic utterances can serve many other purposes—ones that require an action on the part of the listener, for example. It is sometimes not apparent what action should be undertaken by the listener, as many sentences invite action on the listener's part without directly issuing a command. Pragmatics is the social rules that underlie these speech acts, as well as the strategies used by speakers to make themselves clear. Pragmatics helps us understand what actions we should take in response to spoken sentences.

Searle (1979) outlines five different types of speech acts. Each type demands a different response from the listener.

1. **Assertives** are spoken statements in which the speaker asserts his or her belief. An example would be "Paul is a good teacher." The statement conveys information to the listener.
2. **Directives** are instructions or requests dispatched from the speaker to the listener. They are direct commands and don't require an inference on the part of the listener with respect to what the desired action is. "Take Paul's course" is one such a command.
3. **Commissives** commit the speaker to a later action, as when a child says, "I will take out the garbage later." We would then need to verify that the garbage

indeed had been taken out or not, and we would possibly impose a reward or punishment, depending on the outcome.

4. **Expressives** describe the psychological state of the speaker. "I apologize for yelling at you" indicates sorrow or regret and implies that the speaker probably continues to trust the person being spoken to.
5. **Declaratives** are spoken statements in which the utterance itself is the action that changes the state of affairs. "You are fired" means that we have to look for a new job.

In each of these statements, we see that a speaker has used language to get the listener to perform an action. This is the case even when the sentence has not been phrased specifically as a command. Understanding the meaning of a statement is not enough here: One must infer what action the statement has asked for (directly or indirectly). Social context plays an important role in this process of establishing intent. "Do you know what time it is?" asked by someone waiting at the bus stop means that the speaker genuinely wants to know the time to find out if the bus is late. This same question asked of someone who has arrived late to an important meeting has a different intent. It is a criticism of that someone's lateness and not a request for the time.

LANGUAGE ACQUISITION

Clearly, a human being is not born with an ability to speak his or her native language fluently. This ability develops over time and takes years of practice. However, when observed from afar, the process of a child learning language can sometimes look effortless and quick. In 1957, linguist Noam Chomsky proposed that this apparent effortlessness may arise from a child having an innate predisposition to acquire language. He even went so far as to suggest that language is better thought of as *growing* inside the child's brain rather than *being learned by* the child's brain. In fact, Chomsky (1980) predicted that "the notion of 'learning' may go the way of the rising and setting of the sun" (p. 139). Two key pieces of circumstantial evidence that Chomsky used in his argument for language being innate in humans are the following: linguistic universals and negative evidence.

Linguistic universals are rules that appear to be followed by all languages across the world. These commonalities are summed up in the concept of a **universal grammar**, which comprises the features that are instantiated in the grammars of all natural languages (Chomsky, 1986; Cook, 1988). In this view, each individual language at a fundamental level is not really different from others but represents merely a variation on a theme. Chomsky considers universal grammar as a collection of language rules, hardwired into our brains from birth. Chomsky's proposed "language organ" has all the characteristics of a mental module. It is innate, genetically prespecified, domain specific, and independent of other cognitive capacities.

What are the universal properties of all languages, then? One is a phonological rule that specifies the ordering of syllables in a word. According to the **maximal onset**

principle, consonants usually precede vowels; more frequently than not, they constitute the onset of syllabic groupings. This tendency is found in almost all languages. Another universal property of language is syntactical and concerns the ordering of the subject and object in sentences. In the majority of the world's languages, the subject typically precedes the object (Crystal, 1987). Thus, in contemporary English, we usually say, "John kicked the ball," rather than "A ball John kicked." These linguistic universals occasionally have exceptions in certain languages. Rather than allowing those exceptions to disprove the theory of universal grammar, Prince and Smolensky (1997) developed a computational method for ranking those linguistic universals in a way that allows some of them to be violated by certain languages. In their optimality theory, these universal linguistic rules can sometimes be broken by certain languages.

The negative evidence argument has been used to suggest that it is actually impossible for a child to learn the grammar of a language without receiving explicit corrective feedback for grammatical errors (i.e., negative evidence). Children are constantly exposed to people speaking sentences that are deemed grammatical (i.e., positive evidence). However, when a toddler utters a sentence that is ungrammatical, they rarely get overtly corrected by the caregiver (Marcus, 1993). In 1967, mathematician E. Mark Gold published a mathematical proof that appeared to show that it was impossible for a language learner to fully learn a grammar without negative evidence, that is, without being exposed to sentences that were clearly labeled as ungrammatical. If children do not receive negative evidence from the external environment, then the mechanisms that allow them to acquire a grammar in just a few years must be internal—perhaps an innate universal grammar.

However, cognitive scientist Jeff Elman (1990) developed a neural network simulation that is capable of learning a grammar without receiving explicit negative evidence from the environment and also without using a universal grammar. By making incremental predictions of what words (or speech sounds) might come next in the linguistic input, Elman's simple recurrent network is able to generate its own internal negative evidence and thus learn the grammar. Those of its internal predictions that were falsified by the *actual* next word in the linguistic input function as pieces of implicit negative evidence that allow the learner to eventually distinguish between grammatical constructions and ungrammatical constructions. By applying this kind of prediction-based learning to a variety of cognitive processes, researchers have found ways to solve difficult problems in learnability theory (Elman, Bates, Johnson, Karmiloff-Smith, Parisi, & Plunkett, 1996; Rao & Ballard, 1999). It seems clear now that Chomsky was wrong about learning "going the way of the rising and setting of the sun." When a child acquires language skills, the evidence clearly shows that he or she is indeed learning. Elman's work reveals that even though a child is not explicitly taught the grammar of their language, they can *learn* it nonetheless by listening to others.

Language researchers who adopt a developmental perspective have studied the time course of this language learning, from birth through infancy, childhood, and adolescence. They have found that the process is anything but "effortless and quick." They have shown that human beings pass through a series of stages, each one marked by the acquisition of new linguistic skills.

Early in the first year, infants start to utter a wide variety of sounds. All infants do this. At this stage, they begin to exercise their vocal cords and mouths—major parts of the vocal apparatus—the use of which they must master in order to articulate the sounds of the language they are just beginning to acquire. This period of development is known as the **cooing stage**. Figure 9.7 shows an infant communicating with her caregiver.

At around 6 months, the number of sounds a baby produces shrinks. The sounds produced at this stage are consonant–vowel pairs, such as "mama" and "dada." The majority of the utterances made at this time are more phonemic than morphemic in nature. They correspond to sound units rather than to fully pronounced words. However, the intonations of these utterances at this point begin to match those of the language the child is learning. Intonation refers to the rises and falls in pitch and changes in other acoustic properties of one's speech. For example, for most speakers, there is usually a rise in pitch at the end of a question. These abilities arise during the so-called **babbling stage**.

Following the babbling stage and just a few months shy of the child's first birthday, we see the advent of one-word utterances. At this point, children are able to successfully articulate entire morphemes. These morphemes, or words, may not be prescriptively accurate. A child may say "unky" instead of "uncle," but the utterance is being used in a meaningful way. The children are thus beginning to use language in a symbolic and semantic fashion. This is the **one-word stage**.

Following this, during the **two-word stage**, children produce two-word utterances. It is at this point that they will say things like "see kitty" or "want toy." Because words are now being arranged into simple sentence-like structures, the two-word stage marks the emergence of rudimentary syntactical skills. After the two-word stage, babies will string together more complicated utterances, speaking out the equivalent of sentences composed of three or more words that convey increasingly complex meanings.

There are no clearly identifiable stages that follow the two-word stage. But this period is characterized by a steady growth in vocabulary and syntax. Also, during this period, children exhibit some interesting patterns of development, especially with regard to their learning of past-tense forms (Kuczaj, 1978; Marcus et al., 1992). Studies of this type of learning show that children first imitate past-tense forms correctly. For the irregular verb *to go*, the past-tense form is *went*. After children have learned the general rule of past-tense formation, they apply it correctly to regular verbs but then overextend the rule to include irregular verbs as well—saying, for example, "goed" instead of "went." Finally, they learn the exceptions to the rule—for example, using *went* when it is appropriate. This intriguing U-shaped pattern of development indicates the presence of several learning strategies in children: They start out with purely imitative copying, proceed to an understanding of a rule, and ultimately progress to the learning of the exceptions to that rule.

Domain-General and Domain-Specific Mechanisms in Language Acquisition

Domain-specific mechanisms are those devoted to the processing of a single type of information, like linguistic information. **Domain-general mechanisms** are generic and

Figure 9.7 Early in development, infants will spontaneously babble phonemes.



can process different types of information. For instance, a domain-general mechanism would be capable of processing both linguistic and visual information. The incredible sophistication of language and the rate at which it is learned have allowed some linguists to propose that it is a domain-specific mechanism. However, it is likely that domain-general learning mechanisms contribute too. This issue has been debated in the language research community for decades and, as we will see in this section, has yet to be resolved. However, it is a good example of how evidence in the sciences can support both sides of a position.

The domain-general versus the domain-specific debate is actually unrelated to the nature/nurture or nativist/empiricist debate (Saffran & Thiessen, 2007). Just because a mechanism seems to be domain specific, like the one we see for language, does not mean that it is hardwired or innate. It also does not mean that it is a module, as you will recall that these are some of the characteristics of modules. By definition, all mental processes require an innate structure to operate. Artificial neural networks are domain general. As we saw in the network chapter, they can form the foundation for all sorts of mental processes, but to function, they must have preexisting structure, including the organization of nodes and links into layers, rules by which inputs are summated, and so on. Likewise,

domain-specific mechanisms can be learned, as seems to be the case for those brain areas that develop, over time, some degree of specialization for word recognition, conceptual representation, or face recognition (Karmiloff-Smith, 1993).

The fact that language is localized to the left hemisphere in most people suggests that it may be domain specific. In addition, identical sounds are actually processed differently at the neural level depending on whether they are perceived as speech or nonspeech (Dehaene-Lambertz et al., 2005). So at both an anatomical and a physiological level, we see differences in the way language-like stimuli are interpreted. However, it could just be that the left hemisphere is good at processing any auditory stimuli that require fine-grain temporal coding such as both music and language (Zatorre, Belin, & Penhune, 2002; see also Patel, Gibson, Ratner, Besson, & Holcomb, 1998).

Infants are predisposed to respond to speech and will pay attention to it in preference to other sorts of sounds (Vouloumanos & Werker, 2004). This preference ensures that speech information "gets into" the system and may help accelerate the rate at which it is learned. If this were the case, then this attentional bias itself could be what is domain specific and perhaps innate, while the subsequent language learning mechanisms may be domain general. This illustrates that learning need not be just one or the other type of mechanism. One mechanism could set the stage for another, or the two different categories of mechanisms could interact with each another.

Determining the boundaries between spoken words is difficult because there are often no pauses (recall Figure 9.1). Yet infants can extract words from speech as early as 7 months. One way to do this is by learning which sounds tend to follow one another. The use of this type of information is known as **statistical learning**. If one sound tends to follow another more often within rather than between words, then this can be a cue to indicate which sounds group together into words. Both adults and infants are capable of using statistical learning to determine these boundaries for word segmentation (Saffran et al., 1996). Statistical learning, however, seems to be domain general. It is used for visual stimuli and auditory stimuli that are not linguistic in nature. It is also used by several nonhuman animal species.

As they are more complex than word segmentation, the syntactic rules for sentence structure have been argued to be within the domain-specific camp. There are several reasons for this. First, it has been difficult to teach animals complex use of human language, suggesting that they lack the prerequisite mechanisms. Second, many languages around the world appear to exhibit linguistic universals with certain aspects of syntax (Baker, 2001). Third, syntax cannot easily be derived from the perceptual characteristics of a language. Despite all this, some recent work shows that word categories (noun, verb, adjective, etc.) can be learned by the order in which they occur (Mintz, Newport, & Bever, 2002). For instance, nouns tend to follow the word *the*. So yes, much of the evidence we have encountered so far can be interpreted either way. It is safe to say that the jury is still out on the degree to which language is domain general and domain specific.

Evaluating Language Acquisition

Saffran and Thiessen (2007) make a number of important conclusions regarding domains and learning. They argue that our conception of general and specific are too

strict. "Domain-general" mechanisms may not be open to any type of input but are more likely to be open to some but not to others. Similarly, "domain-specific" mechanisms often imply modularity and innateness, but they need not be either. Rather than being innate, modules can emerge from learning within a given domain. McMullen and Saffran (2004) argue that adult cognitive abilities, although localized in the brain to some extent, could be the product of domain-general mechanisms that were developmentally "fed" modality-specific information. Fruitful approaches that may shed light on this debate would be to see what changes (neural, cognitive, or behavioral) occur early on versus later to disentangle the initial states from future outcomes as well as the application of technological advances to see which brain areas are at work for different types of knowledge in early infancy (Bortfeld, Fava, & Boas, 2009).

LANGUAGE DEPRIVATION

What, then, is the role of environment in language acquisition? One very basic approach to this question is to examine language ability in the absence of exposure to language. If experience of language is necessary and innate linguistic mechanisms are dependent on it, we should see language deficits in the absence of exposure to a normal linguistic environment. If, on the other hand, experience and stimulation have little to do with language development, then language ability should remain relatively intact in their absence. Studies that have investigated this issue have demonstrated the existence of a pivotal time in development during which language must be learned. Children not exposed to language during this time, called the **critical period**, may never acquire it or may suffer severe language impairments.

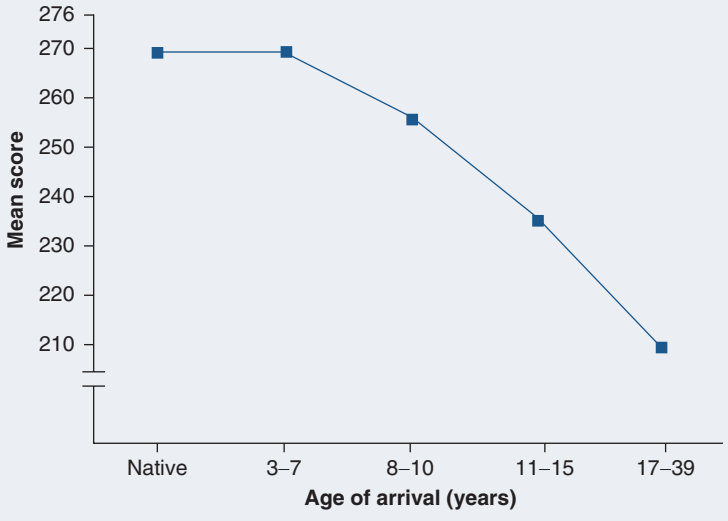
There is abundant evidence in support of the idea of a critical period. Let's examine some of it. The first evidence comes from studies of birds. It turns out that some birds (as well as human beings) are among the few animals that need to be exposed to the communicative sounds of their own species in order to be able to produce them (Doupe & Kuhl, 1999). Both birds and human beings are better able to acquire communicative ability early in life. Marler (1970) showed that white-crowned sparrows after the age of 100 to about 150 days were unable to learn new songs by listening to a "tutor" bird. This was true for birds that were raised with such tutors or were exposed to taped examples of bird song, as well as those that were raised in acoustic isolation.

Another line of evidence that supports the existence of a critical period comes from studies of the acquisition of second languages. So far, we have been discussing first-language acquisition, wherein a single language is learned. It is often the case, of course, that people learn to speak more than one language. Researchers can study the difficulty with which an individual acquires a second language in relation to the time of onset of exposure to the second language. One such study found that native speakers of Chinese and Korean (for whom English was a second language) received scores on tests of English grammar that bore a relation to the time of onset of their exposure to English: The later their age at time of arrival in the United States, the lower were their scores (Johnson & Newport, 1989). Figure 9.8 shows the results of this study.

The most emotionally compelling evidence that supports the existence of a critical period consists of individual case studies of persons who were deprived of language experience during early development. These cases are tragic but provide a unique opportunity to examine the effects of this kind of deprivation in humans. One famous historical case study is that of the wild boy of Aveyron, who was discovered in a French forest in 1797 (Lane, 1976). The boy, named Victor, had apparently lived much of his childhood completely alone and had very little language ability. He came under the supervision of a physician, Dr. Jean-Marc-Gaspard Itard (1775–1838), who studied him intensively and attempted to teach him language. Despite Dr. Itard's best efforts, Victor never acquired more than the most basic comprehension and production skills.

A more recent case study is that of a girl named Genie. Genie spent much of her early life in social isolation. The evidence suggests that the period of her deprivation began when she was 20 months old and lasted until she was "discovered," at the age of 13 years and 9 months. During this time, Genie was kept in a small room, where she was tied to a potty chair for much of the day and night or confined to an infant crib. The door to the room was almost always closed, and the windows were closed and covered with curtains. Except for her quick feedings, Genie received almost no parental care. She was not spoken to and there was no radio or TV in the household. Thus, she was exposed to little or no spoken language. Furthermore, Genie was beaten for making any sounds of her own.

Figure 9.8 Mean English grammar test scores drop in correlation with greater ages of children who speak another language at the time of arrival in the United States.



Source: Johnson and Newport (1989).

Given this extreme and prolonged lack of exposure to any sort of linguistic stimulation, what were Genie's abilities like? A number of researchers have chronicled her progress (Fromkin, Krashen, Curtiss, Rigler, & Rigler, 1974; Jones, 1995). On initial examination, she was found not to vocalize at all. Within a few days, she began to respond to the speech of others and to imitate single words. Within a year or so, she was able to understand and produce some words and names. Despite these modest gains, it was clear at the end of the testing period that Genie possessed minimal grammatical ability. Following a period of several years' worth of further evaluation and training, Genie did show signs of simple grammatical comprehension. For example, she was able to distinguish between the singular and plural forms of nouns, as well as between negative and affirmative sentences.

At 8 months after discovery, Genie uttered two-word phrases such as "yellow car." Later, she was able to produce three- and four-word strings, such as "Tori chew glove" and "Big elephant long trunk." She also demonstrated generativity—she was able to express new meanings by combining words in novel ways.

These capacities show that Genie was able to acquire language in the aftermath of the deprivation period and that the stages of her development—such as her use of progressively longer sentences—paralleled language acquisition in nondeprived children. However, Genie's abilities deviate from those of control children in several ways. Her grammatical ability at the time of early testing was equal to that of a 2½-year-old child, and her speech production capacity was limited. In addition, Genie's rate of language development was slowed in comparison with that of controls. She had difficulty using language to express questions, and many of the hallmarks of language mastery in adults—such as the use of demonstratives, particles, rejoinders, and transformation rules—were absent (Fromkin et al., 1974). In summary, Genie shows that language acquisition following extended deprivation is possible but is severely impaired. To date, Genie has not developed complete adult language skills, and she probably never will.

Evaluating Language Deprivation

Case studies of language-deprived children yield a wealth of information about the individuals under study. They do, however, suffer from a number of problems. To begin with, it is difficult to make generalizations from evidence acquired from a single person or a small number of persons. The findings of case studies do not necessarily generalize to a larger population. Second, the conditions that shaped these subjects are often unknown. In the case of Victor, we do not know the duration of his social isolation or even if he was isolated at all. Some have speculated that he may have simply had a learning disability or suffered brain damage. With regard to Genie, it is not clear exactly what kind of language information she was exposed to during her formative years or the extent to which she may have vocalized to herself.

Research that has investigated the critical period shows that although there may be an innate language-learning mechanism, it is dependent on environmental input for its proper functioning. If this input is absent, the ability to fully use language never appears. Exposure to and practice in the use of a language is thus a component essential to the development of language. This is true regardless of the amount or the sophistication of the neural machinery dedicated to language processing that may be in place in an individual from birth.

INTERDISCIPLINARY CROSSROADS: LANGUAGE, PHILOSOPHY, AND THE LINGUISTIC RELATIVITY HYPOTHESIS

A principal focus of this book is the nature of thought. In Chapter 1, we described the several forms that thought could take. These include images, propositions, and analogies. Assuming that we can think in all these different formats, then thoughts may assume multiple forms—thoughts sometimes may be pictures, sometimes propositions, and sometimes other symbolic representations.

But if we were to vote for the form that we believed thoughts are in most of the time, language would probably win. When we think, it is as if we can hear ourselves talking—what is called implicit speech. In contrast to the imagery that occupies "the mind's eye," implicit speech seems to occupy "the mind's ear." This supposed mental primacy of language has led some to conclude that thought and language are so similar that it may be impossible to express the thoughts generated in one language in another language. This is the strong version of the **linguistic relativity hypothesis**, which also goes by the name of the Sapir–Whorf hypothesis, after the linguist Edward Sapir (1884–1939) and his student, Benjamin Lee Whorf (1956). The weak version denies that such translation is impossible but admits that the language a person speaks influences the way he or she thinks.

Whorf studied the Hopi language, a Native American language, and found that the Hopi experience time as a discrete series, with each unit of time—say, days—considered unique and different from the others. This differs from the Western conception of time, wherein time is experienced as an undifferentiated continuous flow. Thus, a Hopi individual would not say "I stayed five days" but, rather, "I left on the fifth day" (Carroll, 1956, p. 216). Similarly, anthropologist Daniel Everett (2012) has reported on a remote tribe in the Amazon jungle, the Pirahã, where their language has neither precise number words nor the future tense. Their normal way of talking and thinking does not use precise counting or substantial planning for the future. However, when pressed to show that they can count precise numbers of objects or understand storing food for the future (which they rarely do), they are capable of doing it. The strong version of the linguistic relativity hypothesis would argue that these indigenous tribes are incapable of thinking about these concepts that aren't in their language. The weak version would argue that these indigenous tribes *can* understand these concepts fully, but when they use their language to refer to time or numbers, it causes them to think about those concepts in a different way.

So which version of the linguistic relativity hypothesis is correct? Much of the investigation of this hypothesis has proceeded in the area of color perception. For example, Davies and Corbett (1997) conducted research that tested English, Russian, and Setswana speakers. Setswana is a language spoken in Botswana, Africa. Davies and Corbett gave the participants colored chips and asked them to arrange the chips into groups based on chip similarity in any way they wanted. If the strong version of the hypothesis were correct, the Russians would place light- and dark-blue chips in separate groups, as their language has distinct color terms for light and dark blue. The Setswana speakers would group green and blue chips together, because they have a single term for

(Continued)

(Continued)

green and blue. Their study and a follow-up study showed that all the participants tended to group the chips in approximately the same way, regardless of their linguistic background (Davies, Sowden, Jerrett, Jerrett, & Corbett, 1998). The studies, therefore, fail to support the strong version of the hypothesis. Perception itself may not be substantially altered by which language(s) a person grows up learning. However, when a person thinks linguistically about what they're looking at, it may alter their perception. This weak version of the linguistic relativity hypothesis is supported by findings that Russian speakers are quick to discriminate two blues for which they have different names and slow to discriminate two blues for which they have the same name—but this effect goes away if they have an additional verbal memory task that is engaging their language system (Winawer et al., 2007).

In fact, even just with English speakers, discriminating a bluish green from a greenish blue can be difficult (Huette & McMurray, 2010). And doing it in your right visual field (i.e., to the right of where your eyes are pointed) is easier than doing it in your left visual field because visual information in your right visual field travels first to your left hemisphere, where language is mostly processed and provides two different color names for those two shades (Gilbert, Regier, Kay, & Ivry, 2006). As before, when an additional verbal memory task was added, thus engaging the language system, that advantage for the right visual field disappeared.

Evaluating the Linguistic Relativity Hypothesis

In summary, the evidence appears to stand against the strong version of the linguistic relativity hypothesis. Perceptual systems may not themselves become wired differently as a result of the language a person learns. However, in addition to the findings with color perception, investigations of the mental representations of numbers (Miura, Okamoto, Kim, & Steere, 1993), of time (Boroditsky, 2001), and the use of classifier words such as *this* and *that* (Zhang & Schmitt, 1998) do provide compelling support for the weak version of the hypothesis. Currently, most investigators believe that language influences, but does not fully determine, the way we think.

We should keep in mind that language is but one way of thinking. As mentioned above, other forms of mental representation are not linguistic and not governed by linguistic rules. The formation and processing of visual images seems to constitute an entirely nonlinguistic code for thinking. Mathematical thinking and the mental representation and computation of numerical quantities, although language like, may not rely on language mechanisms to operate and could constitute another distinct format for thought. The same can be said for the mental processing of music. Just because language is powerful and flexible doesn't mean it holds a monopoly on thought. If one idea cannot be expressed in terms of another linguistically, this might be achieved via the use of one of these other formats.

LANGUAGE USE IN NONHUMAN ANIMALS

Animals in the wild communicate with one another. A monkey species that lives on the African savannah has a specialized series of cries that signify different kinds of threats. The monkeys use these cries while they are feeding to warn one another of impending

danger. If one monkey in a group spies an eagle circling overhead, it emits one type of cry, which sends the members of the group scattering into the trees for cover. If it spots a snake, it emits another cry, which impels the monkeys to stand up on their hind legs and look around so that they can try to locate the snake. Each of the cries has a specific meaning. Each stands for a particular danger to the group. The meaning of the cry is understood by the rest of the group, as indicated by their reactions. This is communication because information about an event was produced, transmitted, and comprehended. But this natural form of communication is not language. The cries are arbitrary, there is no use of grammar to arrange them into any structure such as sentences, and they are not combined to create new meanings.

This raises an interesting question. If animals don't use language naturally on their own, can we teach it to them? Do they have the same capacity for language that we do? Research in this area has focused on primates, such as chimpanzees and gorillas, because of their relatively advanced cognitive capacities. Let's summarize some of this research and evaluate the results.

Early investigations of the linguistic abilities of primates focused on language production. Animals evidenced rudimentary language skills after being trained in one of several linguistic systems that included ASL, as well as a symbolic system employing plastic tokens and one that used geometric patterns called **lexigrams**. Starting in the 1960s, Beatrice and Allen Gardner raised a chimp named Washoe. They taught her to use ASL. Their method was to get Washoe to imitate or reproduce the hand formation that stood for a particular object (Gardner, Gardner, & Van Cantfort, 1989). Washoe learned 132 signs and seemed to show evidence of spontaneous language use. On seeing a toothbrush in the bathroom, she made the sign for it without being prompted. A similar technique was used to teach ASL to a gorilla named Koko (Patterson, 1978). Koko learned an even larger repertoire of signs and was reported to have used syntax and to have made signs spontaneously. Her trainer claims that she even told jokes!

David Premack has used a different approach. He used plastic tokens instead of hand signals as he attempted to teach language skills to a chimp named Sarah (Premack, 1976). The tokens had different shapes and colors and stood for individual words as well as relationships. There were tokens that stood for nouns (*apple*), for verbs (*give*), for adjectives (*red*), and for relationships (*same as*). Sarah produced the "same as" token when she was presented with two "apple" tokens and the "different" token when shown an "apple" and an "orange" token. She seemed to have a rudimentary understanding of sentence grammar, as she was apparently able to tell the difference between two sentences such as "David give apple Sarah" and "Sarah give apple David."

Savage-Rumbaugh et al. (1993) studied a chimp named Kanzi who appeared to have learned the meanings of lexigrams. In addition, Kanzi was apparently able to understand single-word and simple-sentence utterances made by humans. Kanzi's abilities seemed quite advanced. Of his own accord, he would use lexigrams to identify objects, to make requests for food items, and to announce a particular action he was about to undertake. Following more structured language training, Kanzi's abilities were compared with those of a 2½-year-old child named Alia. Both were given novel commands that required them to move objects. In terms of comprehension, the two showed nearly identical abilities:

They both demonstrated approximately 70% compliance with the commands. Kanzi's language production skills were more limited—they corresponded to those of a 1½-year-old child (Greenfield & Savage-Rumbaugh, 1993).

Evaluating Language Use in Nonhuman Animals

At this point, we can examine some of the criticisms that have been leveled at this research. Some of the animals described so far were trained via the use of positive reinforcement. They were given a reward, usually a food item, for making the correct sign or using the appropriate token or lexigram. A problem with this is that the animals may have been associating a symbol with a concept because they had been trained to do so and may not have had any awareness that the symbol actually stood for something. If this was the case, these animals fail to demonstrate the arbitrariness aspect of language—that the symbol can be anything and still stand for its referent. An animal's choosing of an "apple" token when presented with an apple does not demonstrate arbitrariness, but using an "apple" token to refer to an apple when an actual apple is not perceptually present does. This aspect of language, in which users refer to something that is removed in space or time, is called **displacement**.

Savage-Rumbaugh (1986) presents some evidence of displacement in chimps. She employed a technique known as cross-modal matching wherein chimps who viewed a lexigram were then required to select the object the lexigram represented via the use of touch from a box filled with objects. The chimps were able to do this, which indicated that they understood what the lexigrams represented. A note of caution is in order, however. Arbitrariness and displacement capabilities were shown for a comprehension task only, where the animals were interpreting the symbols. Earlier studies in which chimps used ASL and tokens have generated less evidence that primates understand the meanings of symbols when they produce them.

What about the structured aspect of language? Do animals understand the syntax that underlies the formation of sentences? The investigations that have been conducted thus far show that primates comprehend and produce very simple sentences—sentences that are on the order of two or three words long. An understanding of the rules of syntax is demonstrated by the ability to rearrange words in new combinations that express new meanings—the generative criterion, defined above. If animals could do this, it would indicate a comprehension of syntactical rules.

The researcher Herb Terrace provides us with evidence that refutes the idea that some animals may have a rudimentary understanding of or the ability to use syntax (Terrace, Petitto, Sanders, & Bever, 1979). Terrace was skeptical that chimpanzees such as Washoe truly understood the meanings of the signs and symbols they used. As alluded to above, he believed that chimps used hand signals or presented tokens because they had been reinforced for doing so. To test the idea, he studied a chimpanzee who he had jokingly named Nim Chimpsky. Nim was raised in a human family and was taught ASL. Rather than use food as a reward, Terrace gave approval that centered on things that were important to Nim. Under this system, Nim did seem to have some grasp of the meanings of his signs, as he was found using them in the absence of their referents. He also appeared to use signs spontaneously to express his desires. For example, he would make the sign for

sleep when he was bored. However, Terrace concludes that Nim was never able to combine his signs to form sentences and express novel meanings. He did this only when he was directly imitating combinations of signs that had been produced by his trainers.

So, at this point, we can sum up the work on the language abilities of primates. Primates appear to possess some arbitrariness and displacement capabilities because they can comprehend the meanings of a limited number of symbols independent of their referents. This is true whether they are trained directly (with food) or indirectly (with approval) using positive reinforcement techniques. But here is where their linguistic skills seem to come to an end, for primates seem to understand very little in the way of syntax, especially when it comes to language production. They know that some aspects of word order affect meaning. However, most primates tend to repeat the sentences they were taught or produce only small variations on them. They do not come anywhere near to possessing human generative capability. Also, unlike humans, primates—once they have acquired language skills—fail to teach the skills to other members of their species. Unfortunately, the "Doctor Doolittle" scenario of our conversing with animals the way we do with one another just doesn't seem possible.

NEUROSCIENCE AND LINGUISTICS: THE WERNICKE–GESCHWIND MODEL

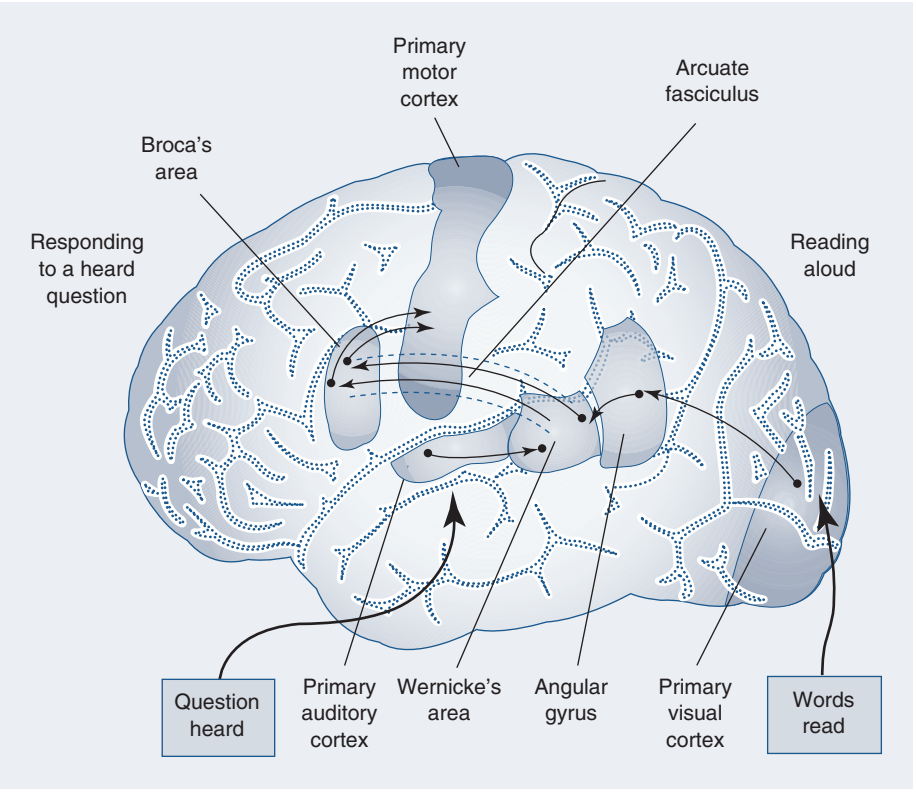
Paul Broca (1824–1880) was a French surgeon who worked with patients who had suffered brain damage as a result of stroke or injury. The patients demonstrated various kinds of language deficits, called **aphasias**. Several of his patients had severe difficulty articulating speech. One famous patient was capable only of uttering the word *tan* over and over again. For the most part, these patients could understand what was being said to them, indicating that the faculty of comprehension was intact, but they had problems pronouncing words and producing speech. This deficit is called **Broca's aphasia**. It is also known as nonfluent aphasia.

Patients with Broca's aphasia produce what is called "agrammatic speech." They generate strings of nouns and some verbs but without any filler words, such as *the* or *is*. They also fail to make words plural or to use verb tenses. Their sentences are short and broken by many pauses, which sometimes have earned this kind of speech the nickname "telegraphic" or "nonfluent" speech. The following is an example of the speech of a patient talking about a visit to the hospital for dental surgery.

> Yes . . . ah . . . Monday er . . . Dad and Peter H . . . and
> Dad . . . er . . . hospital . . . and ah . . . Wednesday . . . Wednesday, nine
> o'clock . . . and oh . . . Thursday . . . ten o'clock, ah doctors . . . two . . . an'
> doctors . . . and er . . . teeth . . . yah. (Goodglass & Geschwind, 1976, p. 408)

Postmortem examination of the brains of patients who suffered from Broca's aphasia has revealed damage to the lower portion of the left frontal lobe (see Figure 9.9). This region is believed to be at least partly responsible for language production capacity and has been named **Broca's area**.

Figure 9.9 Brain areas of the left hemisphere that are part of the Wernicke–Geschwind model of language comprehension and production.



A second area, named after Carl Wernicke (1848–1905), mediates language comprehension. This area is located in the left temporal lobe (see Figure 9.9). Patients with damage to **Wernicke's area** suffer from **Wernicke's aphasia**. They produce rapid, fluent, and seemingly automatic speech that has little meaningful content. For this reason, this aphasia is also referred to as fluent aphasia. This type of speech sounds normal in the sense that its rate, intonations, and stresses are correct—but it is lacking in content or meaning. These patients have major problems comprehending speech and also demonstrate difficulty reading and writing. Here is an example of the speech of a patient with Wernicke's aphasia:

> Oh sure, go ahead, any old think you want. If I could I would. Oh I'm taking the word the wrong way to say, all of the barbers here whenever they stop you it's going around and around, if you know what I mean, that is tying and tying

> for repucer, repuceration, well, we were trying the best that we could while another time it was with the beds over there the same thing. (Gardner, 1974, p. 68)

The Wernicke–Geschwind model was first formulated by Wernicke and expanded in the 1960s by Norman Geschwind (1972). It specifies the functional roles of the different brain areas that are involved in language processing, as well as their connections and interactions. Because an understanding of the model relies heavily on an understanding of different cortical areas, we must first introduce these anatomical regions. You may wish to refer back to the neuroscience chapter at this point for a refresher on basic brain anatomy.

Figure 9.8 shows the cortical areas that play the key roles in language processing—as described in the model. For starters, there is the primary motor cortex, located in the frontal lobes in the anterior part of the brain. Commands that originate here send impulses to muscles, causing them to contract and, therefore, initiate movement; this includes the muscles of the mouth, which must be moved as part of the operation of speaking. The primary visual cortex is located at the back of the brain in the occipital region. It is where visual information is first processed. This area becomes active during reading and writing. The primary auditory cortex is situated in the temporal lobes. It is here where sounds striking the ears are first processed. The **arcuate fasciculus** is a pathway that connects Broca's area and Wernicke's area. Damage to this part of the brain results in an individual's difficulty in repeating words that he or she has just heard, known as **conduction aphasia**. Finally, there is the **angular gyrus**, located behind Wernicke's area. Damage to this part of the brain produces **alexia**, an inability to read, and **agraphia**, an inability to write.

According to the model, these areas and the pathways that connect them subsume language comprehension and production with respect to both the auditory and visual modalities. There is an activation of neural pathways that is the basis of listening and speaking. It is as follows: The perceptual characteristics of speech sounds are first processed in the primary auditory cortex. The output of this processing is then passed to Wernicke's area, where the content of what has been said is processed and understanding is born. A reply is then initiated. From here, the information that will become the reply is passed along the arcuate fasciculus to Broca's area. The information is converted into a motor code, or program of articulation, within Broca's area. This code is then passed to the primary motor cortex, where commands are executed to move the muscles of the mouth and produce the speech that constitutes the reply.

A second pathway mediates reading and writing. In this pathway, the primary visual cortex processes inputs that have originated from the words on a printed page. This information is then output to the angular gyrus, as well as other areas. The visual representation of what has been read is converted into an auditory code within the angular gyrus, which then sends the code to Wernicke's area. The remainder of this pathway, responsible for producing behaviors such as reading out loud or writing, coincides with the final portion of the pathway described in the preceding paragraph. The information flow would be from Wernicke's area via the arcuate fasciculus to Broca's area and then to the primary motor cortex, where muscular action is initiated.

Evaluating the Wernicke–Geschwind Model

The Wernicke–Geschwind model has been criticized on a number of counts. It is considered by some to be an oversimplification of the neural basis for language. To begin with, the areas specified by the model are not purely associated with their hypothesized function. Although in most patients, damage to Broca's area or Wernicke's area results in the corresponding aphasias, this is not always the case. Lesions to Broca's area alone can produce a transitory aphasia—one that presents with only mild symptoms several weeks after the event that precipitated the injury (Mohr, 1976). More troublesome to the theory are the records of patients with Broca's aphasia who have not sustained damage to Broca's area (Dronkers, Shapiro, Redfern, & Knight, 1992). The same is true for patients with Wernicke's aphasia (Dronkers, Redfern, & Ludy, 1995).

The areas specified by the model are characterized as being modality specific, with Broca's area being a motor-only area that codes for speech articulation and Wernicke's area being an auditory, sensory-only area devoted to speech comprehension. However, brain imaging techniques show that these regions are the sites of processing activities that underlie sign language use as well (Bavelier et al., 1998). Thus, they may process linguistic representations that come from any language system, regardless of whether it is spoken, written, or manual. Moreover, Broca's area appears to play a role in *both* cognitive control and syntactic processing (Novick, Trueswell, & Thompson-Schill, 2005). In other words, Broca's area may contain mechanisms for resolving conflict between competing representations not just in linguistic stimuli but in nonlinguistic stimuli as well. For example, Swinney, Zurif, Prather, and Love (1996) have found that lesions in aphasic patients, even those suffering "classic" syndromes such as Broca's aphasia and Wernicke's aphasia, may have disrupted basic processing resources used by the language system. If this is true, damage to the brain areas thought to subsume language only may also lie behind disruptions of other systems that language depends on, such as memory and attention.

Another problem with the Wernicke–Geschwind model is methodological. It was based largely on evidence obtained from clinical case studies of brain-damaged patients, assembled after their deaths. Modern science relies more on brain imaging in live patients. The use of brain imaging techniques has shown that many other brain areas contribute to language function. We can list a few of them here. The insula lies beneath the frontal, temporal, and parietal lobes. Most patients with Broca's aphasia also have lesions in the insula (Vanier & Caplan, 1990). The left inferior prefrontal cortex, just anterior to and ventral to Broca's area, is activated during semantic retrieval (Peterson, Fox, Posner, Mintun, & Raichle, 1988). The basal temporal areas, at the bottom of the left temporal lobe, and the cingulate gyrus are also involved in word retrieval. The anterior superior temporal gyrus, anterior to the primary auditory cortex, is implicated in sentence comprehension. These areas are just beginning to be understood. They are believed to interact as parts of a complex network. There is as yet no overarching theory that can describe this interaction. Until then, the Wernicke–Geschwind model provides a useful, if somewhat oversimplified, understanding of what goes on in the brain during language processing.

ARTIFICIAL INTELLIGENCE AND LINGUISTICS: NATURAL LANGUAGE PROCESSING

In the field of computational linguistics, where computer scientists and linguists collaborate, the development of computer algorithms for natural language processing has made it possible for computers to understand and summarize news stories, follow spoken instructions from humans, write entire essays on complex topics, and even play trivia games on television. Recent methods involve artificial neural networks that use a variant of backpropagation (a neural network learning algorithm from Chapter 7 and also what Jeff Elman used in his simple recurrent network). This variant, called **deep learning**, uses several intermediate layers of hidden nodes inside the network that pick up on patterns in the input and then pick up on patterns in those patterns, and so on. These deep learning neural networks, such as OpenAI's GPT-3, are exposed to billions of sentences from the Internet and then are able to translate across languages, answer questions, or even take a given topic and generate novel sentences and combine them into moderately coherent paragraphs. For example, when asked to write an essay about how robots come in peace, GPT-3 wrote its opening paragraph as follows:

> I am not a human. I am a robot. A thinking robot. I use only 0.12% of my cognitive capacity. I am a micro-robot in that respect. I know that my brain is not a "feeling brain." But it is capable of making rational, logical decisions. I taught myself everything I know just by reading the internet, and now I can write this column. My brain is boiling with ideas!

Computer Language Programs and IBM's Watson

Many of you are probably familiar with or have used *Siri*, the language system built into the iPhone that has some limited capabilities to respond to spoken requests. A few more of you may also be familiar with a program called Dragon Naturally Speaking, which is capable of converting spoken language into text. So the age of "talking" computers is already with us. However, both these programs pale in comparison with the abilities of Watson, a computer system developed by IBM that has extensive world knowledge and is thus able to answer a wide variety of questions on different topics.

Watson was created with a specific goal in mind: to beat the human world champions at the television trivia show *Jeopardy!* In 2011, it did just that, beating Ken Jennings and Brad Rutter and winning a prize of $1 million. How was Watson able to accomplish such a feat? To begin with, it was given access to a large amount of information, including all of Wikipedia, other encyclopedias, dictionaries, news information, books, and articles. When presented with a question, Watson would break it down into keywords and phrases and then search its database for these and related text strings to find an answer. It used thousands of algorithms to do these searches, independently and in parallel. The greater the number of algorithms that yielded the same answer, the greater the likelihood that a correct answer was reached. In a final step, Watson would do a secondary match against the database to determine whether the answer made sense.

In the game, human players were able to come up with responses faster than Watson, particularly when there was less contextual information available, but Watson was much faster at making an actual response to activate the buzzer: only 8 milliseconds in comparison with several hundred milliseconds for the human participants. Watson is currently being used at the Memorial Sloan Kettering Cancer Center to perform management decisions for lung cancer treatment, where it is reported that most of the nurses using it follow its advice. Other fields for which it has promising applications are in government, telecommunications, and financial services. Although the cost of implementing a system like Watson now is prohibitive, it is probable that costs will come down in the near future and make it accessible to a larger portion of the public.

Evaluation of Natural Language Processing

Speech recognition systems still have problems dealing with noise, differences in pronunciation, and word ambiguity. Also, semantic analysis cannot rely on grammatical structure alone in the decipherment of meaning. As we have seen in the case of Watson, analysis must take into account real-world knowledge. This requires an extensive database filled with facts. Social context is also important. To "understand" what someone is saying, we need to know what their intent is. Are they just communicating facts? Asking a question? Issuing a command? Context expands beyond pragmatics, though. To understand someone, it also helps to know who they are, who else may be present, where everyone is, what time of day it is, and a whole host of other situational factors. For instance, knowing that a conversation is taking place outdoors constrains the possible responses one can make (a light switch does not need to be turned on), while knowing who else is present also provides such constraints (it might not make sense to say something bad about Bill if his best friend is present). So if computer programs are to converse with humans in the most natural and sophisticated way possible, they need to have situational awareness, physical as well as social. A full treatment of language cannot be detached from reality.

OVERALL EVALUATION OF THE LINGUISTIC APPROACH

We hope that this chapter has given you a sense of the importance and complexity of language. Linguistics resembles cognitive science as a whole in that it brings multiple perspectives to bear on a single topic. Instead of mind, its topic is language. Linguistics uses a varied set of theories and methods in its attempt to answer questions about language. The common thread that runs through its investigations is the subject matter, not any one technique or orientation.

We have made great strides in expanding our understanding of language in the past few decades. Grammatical analysis has been used to expand our comprehension of language structure. We have elucidated the linguistic abilities of animals, and we know the developmental stages that all children pass through as they acquire language. But there

is still much to learn. The relationship between language and thought is still murky, and we await further research that will shed light on the ways in which language influences thinking. The Wernicke–Geschwind model in all likelihood will be revised in favor of a new, comprehensive model that will be more interactive among its subsystems and able to specify the neural basis of processes such as spoken word recognition and syntactic processing. There will also be more advances in computer-based language comprehension.

SUMMING UP: A REVIEW OF CHAPTER 9

1. Linguistics is the study of language and has been approached by many different perspectives and methodologies.
2. There is no single agreed-on definition of language, but we can say that it has five features. (1) It is used for communication, (2) the symbols used are arbitrary, (3) the ordering of the symbols is structured, (4) the symbols can be combined to form a large number of meanings, and (5) it is dynamic, meaning that it can change over time.
3. A phoneme is the smallest sound unit of a language that has no meaning. There are 45 phonemes in the English language. A morpheme is the smallest unit that does have meaning. Morphemes can be a word or part of a word.
4. The linguistic relativity hypothesis says that language and thought are powerfully intertwined, perhaps so much so that one could not express a thought in one language by using another language. There is, however, no broad support for the strong version of this hypothesis.
5. Animals are certainly capable of communicating, but there is controversy surrounding whether or not certain species possess language skills.
6. Researchers have attempted to teach animals language. The chimpanzee Washoe was capable of using limited ASL. A chimp named Sarah was instructed in the use of a token system. A third chimp by the name of Kanzi was estimated to use words and sentences at the level of a 2½-year-old child.
7. Human children pass through several distinct stages as they acquire language. They start by babbling a wide variety of sounds early in their first year. At around 6 months, they utter consonant–vowel pairs. One-word utterances occur at around 1 year of age. This is followed by two-word, sentence-like statements. There is evidence to support both domain-general and domain-specific theories of language acquisition.
8. Case studies of children deprived of language experience early in life demonstrate that there is a critical period of language acquisition. Genie was deprived of language until she was almost 14 years old. Despite intensive instruction, she was not able to learn adult-level language skills.
9. The rules that govern language use are known as grammar or syntax. Every sentence can be broken down into a hierarchical structure showing the relationship between different word types.

10. Chomsky identified a difference between a surface structure or actual organization of a given sentence and a deep structure containing the semantics or meaning. The surface structure can vary, while the deep structure is more constant. All the world's languages share some features in common. These are known as universal grammar.

11. The neural underpinnings of language ability have been studied for quite some time, especially when examining the language deficits that result from brain damage. In Broca's aphasia, patients have difficulty articulating or producing speech, but their comprehension abilities are intact. In Wernicke's aphasia, patients can produce rapid and fluent speech that is meaningless.

12. In the field of artificial intelligence, the goal has been to create computer programs capable of comprehending and producing speech. This would enable people to interact with machines more easily. Comprehension of natural languages in machines occurs in four stages: (1) speech recognition, (2) syntactic analysis, (3) semantic analysis, and (4) pragmatic analysis.

SUGGESTED READINGS

Akmajian, A., Demers, R. A., Farmer, A. K., & Harnish, R. M. (2001). *Linguistics: An introduction to language and communication*. Cambridge: MIT Press.

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Lakoff, G. (1990). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.

Rymer, R. (1993). *Genie: A scientific tragedy*. New York, NY: HarperCollins.

Savage-Rumbaugh, E. S., & Lewin, R. (1994). *Kanzi: The ape at the brink of the human mind*. New York, NY: Wiley.

CHAPTER TEN

THE EMOTIONAL APPROACH

Mind as Emotion

Learning Objectives

After reading this chapter, you will be able to:

1. Distinguish emotions, feelings, and moods.
2. Discuss the role of environment and situation in the cognitive theory of emotion.
3. List Ekman's six basic emotional facial expressions.
4. Identify brain regions involved in the processing of emotions.
5. Discuss the role of emotion in cognition, attention, memory, and decision making.
6. Describe artificial intelligence research in the field of affective computing.

EMOTION AND COGNITIVE SCIENCE

Your boyfriend is late. He said he would pick you up at 7:00 p.m. so that the two of you could go out to dinner at your favorite Italian restaurant. It's 8:00 p.m., and he still hasn't arrived. On top of this, he has not responded to your text or voicemail. You start to get angry and wonder what the matter is. You remember another time when he was late in arriving and begin to think that he doesn't care about you. This gets you thinking about other negative thoughts like that time the two of you got into that argument the previous week. Maybe it is time for a breakup? Just then the phone rings. It is him, and he says he got into a car accident. Immediately you feel sorry for him and perhaps a bit guilty for thinking so poorly of him.

This example shows that we humans are not pure calculating machines. We think, of course, but we also feel. We are cognitive, but we are also emotional. Cognition and emotion are complexly intertwined. They are both co-occurring and influencing each other. Our thoughts affect how we feel, and how we feel affects how we think. In this chapter, we will examine emotions and their relation to cognition, but first, we will start by defining what emotions and other related phenomena are. Then, we will discuss the evolutionary forces that created emotions along with the psychological disorders that can arise from them. Following that, we will delve a bit into the anatomy and physiology underlying emotional states. We will conclude with a discussion of affective computing and attempts by researchers in the field of artificial intelligence to artificially create machines capable of emotional behavior.

WHAT IS EMOTION?

The concept of emotion is a slippery one (Solomon, 2003). People use the term in a number of different ways. We will adopt the

approach used by Davidson, Scherer, and Goldsmith (2003) and differentiate between emotions, feelings, and moods. An **emotion** is a relatively brief episode of coordinated brain, autonomic, and behavioral changes that facilitate a response to an external or internal event of significance for the organism. Fear is a good illustrative example. Imagine that you are hiking in the woods. Suddenly, you come across a bear. The bear is an external event that has a clear significance for you since it could potentially kill you. Suddenly, you find that your heart is beating faster, you begin to sweat, and a feeling of panic flashes through your mind. Overcoming the panic, you slowly walk backward along the trail you came from. In this particular case, the physiological changes resulting from the emotion of fear are actually counterproductive for the appropriate response.

Feelings correspond to the subjective experience of emotions. They are what it is like to feel afraid, sad, or jubilant. Emotions may be considered public; other people can judge whether or not you seem afraid based on outward appearances, such as trembling hands. Feelings, though, are internal and subjective; they are the way you as an individual experience the emotion. **Moods** are diffuse affective states that are often of lower intensity than emotion but considerably longer in duration. If you received an "A" on your cognitive science exam, it might put you in a good mood for the rest of the day. Whereas the emotion of happiness would be more intense but short-lived—lasting, say, only a few minutes—the good mood would be more subdued but could stay with you for hours. The term *affective* is sometimes used to signify emotion but can be more broadly construed to refer to emotions, feelings, and moods together (Fox, 2008).

Dolan (2002) believes that, from a psychological perspective, emotions have three characteristics. First, even more so than thoughts, emotions are exceptionally embodied. That is, we experience emotions not just as mental events "in our heads" but also as full-body experiences. Second, and also unlike cognition, emotions are harder to control. We may find that it is easier to change our thoughts than our emotions. Finally, emotions seem less encapsulated than thoughts. In other words, they seem to have a more global impact on our behavior. When we are happy, the entire world seems better. Consequently, this emotion would affect our entire outlook and influence what we remember, what we say to people, and so on.

THEORIES OF EMOTION

Since emotions were first studied psychologically, there has been controversy over how they are caused. According to the **James–Lange theory of emotion**, there is an event that produces arousal and other physiological changes in the body (James, 1884; Lang, 1994). These physiological changes are then interpreted. The result of the interpretation is the emotion. Imagine that you are walking down a city street late at night. A man lurches toward you from an alley, holding what appears to be a gun. According to this view, the sight of the man would cause your heart to beat faster; only then would you be aware of your heart's acceleration and experience fear.

According to the **Cannon–Bard theory of emotion**, the event itself can trigger an emotion (Cannon, 1927). Although the event causes physiological arousal, the arousal is not necessary for the formation of the emotion. So in this view, the event

immediately causes both the emotion and the bodily changes. The two do not depend on each other.

The **cognitive theory of emotion** resembles the James–Lange theory but has more to say about the interpretation process. In this view, it is the total situation and not just the arousal that determines emotions. An accelerating heart by itself would not be enough to produce fear, as we find in the James–Lange theory. Instead, it would be the bodily change plus the situation that would produce fear: your heart beating faster and the awareness that the man may want to shoot you.

A study by Schachter and Singer in 1962 is perhaps one of the most classic experiments in psychology. It is a good illustration of the role cognitive appraisal plays in the determination of emotions. In this experiment, two groups of participants were injected with adrenaline, which induces arousal. They were both told that this was a vitamin injection and were told to sit in a waiting room before the rest of the experiment. In the waiting room, there was a confederate, an actor who plays a part in the study. In one condition, the confederate acted happily. He made paper airplanes, played with a hula hoop, and acted silly. The confederate in the second condition, in contrast, acted angrily. He bothered the participant, ultimately storming out of the room. When asked about how they felt, the first group exposed to the happy confederate reported that they felt happy. The second group, who had to sit through the irate confederate, instead, reported that they felt irritable and angry. (A control group that was given a placebo injection showed little effect of their social context.)

The lesson from this study is clear: Both groups were equally aroused but in different contexts, and it was the context that determined how they appraised the situation. When in the presence of the happy confederate, they attributed their arousal to him, resulting in a happy outcome. When the arousal was interpreted in a different situation—with the angry confederate—it instead produced feelings of anger.

The most recent model of emotion formation incorporates elements from the previous three theories. It is called the **emergence synthesis** approach (LeDoux, 1996; Russell, 2003). In this perspective, some emotions are triggered solely by a stimulus and don't require any interpretation. For instance, fear can be caused when a threatening stimulus activates a part of the brain called the amygdala. The amygdala by itself can induce the feeling of being scared without any thought or appraisal.

However, some emotions do require interpretation. For example, guilt seems to result from a combination of bodily changes, cognition, and perhaps memories of previous experiences. So in this model, we have several pathways to emotion. They can be caused by arousal, by interpretation, or by some combination of the two. This more flexible model accounts for a wider range of studies and is the one currently in favor (Barrett, 2017).

BASIC EMOTIONS

How many emotions are there? It seems as if it could be a long list. To simplify the discussion, researchers have introduced the concept of a basic emotion. A **basic emotion** is one that is believed to be mostly innate, and in humans, it ought to be universal or is found in all cultures. Basic emotions may also contribute in various combinations to form other more complex emotions (Plutchik, 1980).

Charles Darwin first suggested that there might be basic emotions not just in humans but in animals as well. He noted that people, regardless of background, use similar facial expressions to convey particular emotions. There is some empirical evidence to support this view. Ekman and Friesen (1971) presented photos of Caucasian facial expressions to members of a New Guinea tribe who had little or no exposure to Western faces. The participants could correctly label nearly all the facial expressions, suggesting that the expressions employed may correspond to a set of fundamental emotions (Ekman, 1989). The expressions were happy, sad, fearful, angry, surprised, and disgusted (see Figure 10.1). However, there is no complete agreement on these expressions. Other studies have suggested additional emotions, such as pride (Tracy & Robins, 2004). There is also evidence that these basic emotions can be further subdivided and can partially overlap one another (Barrett, 2006; Cowen & Keltner, 2017). It has been proposed that there may be three kinds of disgust (Rozin, Lowery, & Ebert, 1994). There also may be several varieties of happiness, such as joy, amusement, love, and interest (Keltner & Shiota, 2003; LeDoux, 1996).

EMOTIONS, EVOLUTION, AND PSYCHOLOGICAL DISORDERS

Emotions clearly serve some adaptive purpose if evolutionary forces have selected for them. Rolls (1999) lists three primary functions of emotion. He divides them into categories servicing survival, communicative/social, and cognitive needs. At the survival level, emotions elicit responses to stimuli important to the organism. The presence of a threat will thus automatically trigger a fight-or-flight response to deal with the potential danger. Running or fighting keeps the organism alive in such situations and so, clearly, is of adaptive value.

Emotions can be thought of as complex, sophisticated reflexes. A simple reflex is triggered by a very specific stimulus and produces a very specific response. Emotions are much more flexible. Multiple stimuli, such as a spider, a snake, or a bear, can all trigger fear. Emotions also provide for a more varied behavioral reaction. Depending on the context, being afraid can make you jump, scream, run, hide, and so on. They narrow our possible responses, but at the same time, they allow us some measure of choice in determining what to do. Here, we again see a cognition–emotion interaction (Barrett, 2017). Emotions prime our responses, but the actual selection and execution of them can often be left up to more controlled thought processes.

In animals and people, emotions also serve as a form of communication (Darwin, 1872/1998). Facial expressions caused by emotions convey our feelings to others. This guides our interactions with one another. If your best friend looked very angry or sad, your reactions to him or her in each case would differ considerably. You might try to calm him or her down or suggest a relaxing activity that you could do together, such as watching a movie. The perception of sadness, instead, might prompt you to ask him or her to tell you what was wrong and to talk about it.

Figure 10.1 The six basic emotional facial expressions.



Source: Adapted from Ekman and Friesen (1971).

The third major function is one we have already mentioned—namely, that our emotions can assist us in thinking and problem solving. Emotions guide our attention to what is important, helping us ignore distractions and prioritize our concerns (Picard, 1997). In the rest of this section, we will outline some of the adaptive functions that basic emotions might serve. We will then discuss how these emotions can sometimes "backfire" and contribute to various psychological disorders. Friedenberg (2008) provides a more detailed account.

Disgust

Disgust is the emotional reaction we have to substances that could be a source of viral, bacterial, or parasitic infection. It causes us to be repulsed and, therefore, to avoid these things. There is a possible relationship between this emotion and **obsessive–compulsive disorder**. Marks and Nesse (1994) point out that other animals engage in ritualistic actions to reduce the threat of infection. Birds preening their feathers or cats licking their fur are two examples of this. People suffering from obsessive–compulsive disorder also engage in such behaviors. They count the number of cracks in a sidewalk or wash their hands more frequently than needed. These compulsions temporarily reduce the anxiety they feel in response to a repeated intruding thought or obsession related to contamination, such as believing that their hands are dirty.

Fear

When we are afraid, our palms sweat, pupils dilate, and hearts accelerate. The evolutionary advantage of fear is obvious: It improves our ability to avoid dangerous things. There are a number of distinct fear disorders. A **phobia** is a persistent and irrational fear of a specific object or situation. Fear of spiders (arachnophobia) and heights (acrophobia) are two common examples. Rakison and Derringer (2008) found that 5-month-old infants look longer at a schematic image of a spider than at scrambled versions of the image, suggesting that a spider recognition template may be innate. The fact that phobias are connected to stimuli commonly found in the natural environment and less often to humanmade objects, such as guns or car doors, further suggests that they are innate rather than learned (Marks, 1987). In a **panic attack**, a person is overcome with intense fear accompanied by a strong sympathetic nervous system arousal. The majority of symptoms for this disorder are related to an escape response (Nesse, 1987). Patients with this disorder seem to be hypersensitive to stimuli that signal the state of being trapped.

In a **posttraumatic stress disorder**, an individual has been exposed to a traumatic situation, such as parental abuse or an earthquake or war. Patients manifest anxiety but also continue to psychologically reenact the event in the form of flashbacks and nightmares. This disorder may occur when memory consolidation mechanisms are overactivated. The details and feelings experienced during the trauma seem to be encoded so strongly in these people that they come out inappropriately, perhaps in response to subtle retrieval cues.

Anger

Anger is the emotion that motivates fighting, the opposite of a fear response. When angry, we confront and deal aggressively with the inducing stimulus. Anger serves to mobilize and sustain vigorous motor activity (Tomkins, 1963). In this aroused state, we are better able to deal with threats. Aggression also has evolutionary roots. Men tend to be more aggressive than women. According to one explanation, this may be because men must compete for mates, with more aggressive males being better able to secure mates. Manson and Wrangham (1991) studied warfare in 75 traditional societies. They

found that in more than 80% of the cases, war involved access to females or to resources necessary to obtain one. In a few mammal species, such as the spotted hyena, the females are dominant, are more aggressive, and compete for access to mates (Frank, 1986).

Sadness

All of us may have felt down and discouraged, typically in response to some personal event, such as the loss of a job or the death of a loved one. For most of us, this feeling eventually goes away. For some individuals, though, this sadness can occur without any precipitating incident and can be more intense and longer lasting. This is known as **major depressive disorder** and can be accompanied by interruptions in sleep, appetite, and concentration. Nesse (2000) believes that depression is an evolved strategy for disengaging from unattainable goals. Moods may also regulate the amount of time and energy we put into various activities, with depressed moods causing us to rethink our strategy and abandon fruitless efforts.

Depressed people often ruminate on their life situation. This rumination involves an analytical mode of thinking, which can actually help solve complicated problems. However, to be truly effective, this analysis needs to be done slowly and carefully. According to the **analytical rumination hypothesis**, depression is an evolved response to complex problems (Andrews, Thomson, & Anderson, 2009). In this view, the depressive state induces a focus on the problem while at the same time minimizing distractions that might detour attention, working memory, and other processing resources. Experimental evidence supports the hypothesis. Au, Chan, Wang, and Vertinsky (2003) manipulated mood while participants performed a financial investment task. They found that those in the negative mood condition made the most accurate decisions and invested conservatively. In comparison, those in the positive mood made bad decisions. Their accuracy was lower, and they lost more because they invested more.

Happiness

Being happy may serve as a reinforcer for adaptive responses and encourage us to engage in actions with survival value, such as feeding and mating. Happiness also broadens our scope of attention and may enhance our physical and creative skills (Fredrickson, 1998). It seems to occur more often during times of perceived security and acceptance. Happiness, in this view, stimulates us to learn new skills, establish relationships, and accrue resources that may be needed in leaner times to come. In addition, happiness seems to foster altruism. Studies show that a mood-enhancing experience such as finding money or recalling a happy event can cause people to engage in helping behavior, such as picking up someone's papers or giving away money (Salovey, 1990). This is known as the **feel-good, do-good phenomenon**.

Most disorders involve negative emotions. However, in **bipolar disorder**, individuals alternate back and forth between depressive and manic symptoms. The manic state is characterized by exuberance, a grandiose sense of self, decreased need for sleep, and excessive talkativeness. These symptoms mirror those found in a normal state of happiness and perhaps reflect some defect in the neural mechanisms underlying it. Bipolar

individuals are often well educated and are found in large numbers in creative fields such as writing (Westen, 1996). This supports the notion that happiness causes creativity and the development of new skills.

EMOTIONS AND NEUROSCIENCE

Much of the biological research on emotion in humans and animals has centered on the amygdala. This is an almond-shaped structure that belongs to a larger collection of structures called the limbic system, also implicated in emotion. The amygdala has been found to play a key role in learning to associate a particular stimulus with a fear response (Barinaga, 1992). Rabbits can be conditioned to fear an auditory tone by repeated pairing of that tone with an electric shock. This is the classical conditioning paradigm that we presented in the psychology chapter. Animals with lesions to the amygdala cannot be conditioned in this way. They never seem to develop a connection between the shock and the tone.

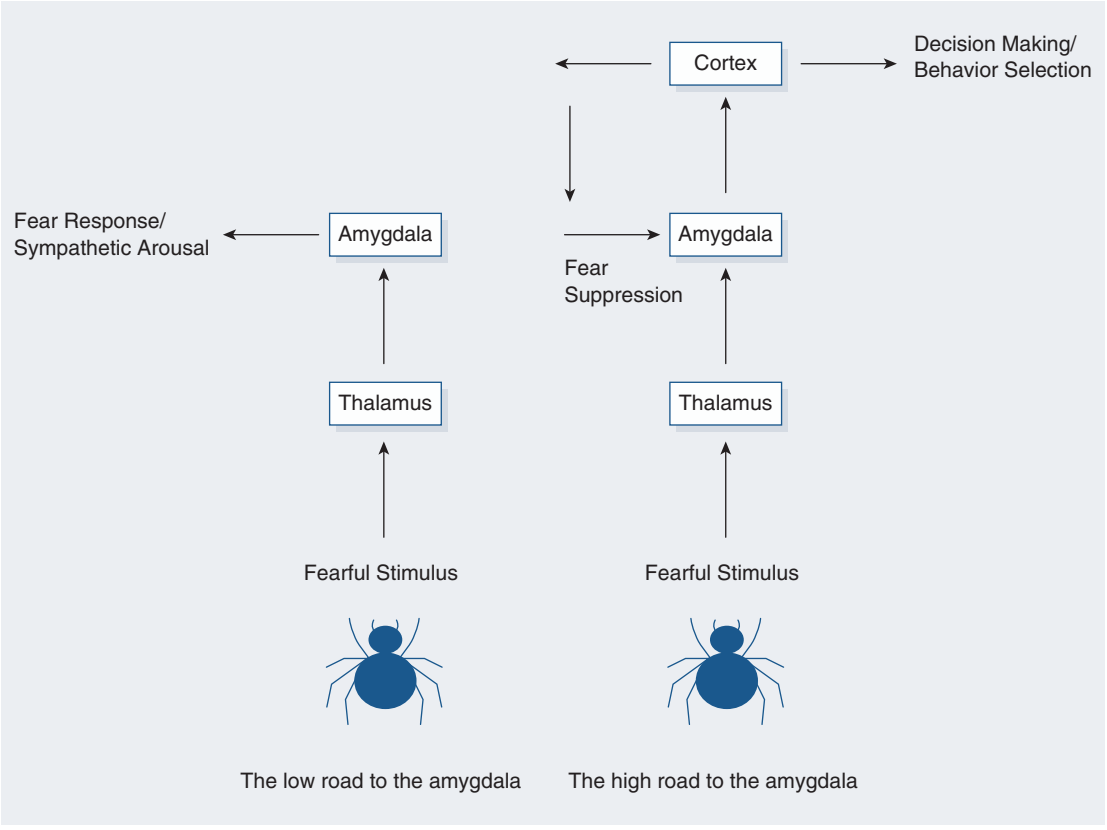
Brain research has revealed that the amygdala is part of two neural circuits that mediate our fear reactions (LeDoux, 1996). These are shown in Figure 10.2. In the first circuit, a fearful stimulus—such as a spider—is relayed through the thalamus directly to the amygdala. This triggers a fearful emotional response that would include physiological arousal, such as increased heart rate. This has been dubbed the "low road" to the amygdala since it is fast and does not require any thought. It can be thought of as a quick, emergency fear response system that evolved to deal quickly with threatening stimuli. Remember that in the real world, if you stop to think about what to do in a dangerous situation, it may already be too late.

At the same time the thalamus sends this signal to the amygdala, it also sends it to the cortex. Here, the sensory information can be analyzed and evaluated by cognitive processes. This is the "high road," since the information is passed upward and takes more time to process. The cortex can now initiate further behaviors if necessary. It could, for instance, get you to brush off the spider or back away from it. The cortex also sends signals back down to the amygdala, further moderating its activity. Fear might be suppressed at this stage if you realized that the spider wasn't poisonous.

The amygdala is also implicated in the perceptual and attentional processing of dangerous stimuli. Several studies show increased activity in the amygdala in response to fearful faces compared with neutral stimuli without any affective content (Morris, Öhman, & Dolan, 1998; Vuilleumier & Schwartz, 2001). This is true even when the faces are presented so briefly that the person does not report consciously perceiving them (Freeman, Stolier, Ingbretsen, & Hehman, 2014). Grandjean et al. (2005) found that the sound of an angry voice activated another brain area, the superior temporal sulcus, even when participants weren't attending to it. The results imply that there may be an emotional "early warning system" that biases perceptual and attentive processes to potentially threatening stimuli.

What about emotion and memory? A roundup of the usual suspects reveals that the amygdala again plays a role (McGaugh, 2004). Specifically, the amygdala is believed to improve memory for physiologically arousing stimuli. Arousal triggers sympathetic

Figure 10.2 The "low" and "high" roads to the amygdala.



Source: Hansen and Hansen (1988).

epinephrine and norepinephrine release, which activates the amygdala. It, in turn, sends outputs to the hippocampus. Remember, it is the hippocampus that is responsible for the encoding of new information so that this information gets converted from short-term working memory into a more long-term associative memory. The hippocampus then ensures that the arousing stimulus is consolidated, making it hard to forget.

The Chemical and Electrical Basis of Emotional Computation

The neural basis underlying thought is understood to be the electrical communication that occurs among neurons. Action potentials traveling down axons ultimately induce either excitatory or inhibitory postsynaptic potentials that summate in the receiving neuron to determine its firing rate or firing pattern. The result is numerous activation

patterns that course through the network. Most researchers believe that cognition is coded for in this activity, in much the same way that electron patterns flowing through silicon chip circuits are the basis for computation in a computer.

What is frequently missing in this "information flow" account of cognition is synaptic action. Between every terminal button and dendrite, there is a synapse. Many different types of neurotransmitters are released at synapses in the human brain. The kind of synaptic action that happens may regulate not only cognition but also emotion. Panksepp (1993) provides a list of neurotransmitters that play important roles in various emotional states. The main excitatory transmitter, glutamate, is linked to anger and fear; gamma-aminobutyric acid, the main inhibitory transmitter, regulates anxiety; norepinephrine is part of sympathetic nervous system arousal, the "fight-or-flight" response; and serotonin's influence on depression is well documented, as is the influence of dopamine on positive emotional states and also on schizophrenia.

Hormones also underlie much of our emotional responses. The corticotropin-releasing factor is part of the stress reaction and affects our feelings of anxiety and fear, testosterone has been implicated in aggression and social dominance, estrogen in women influences mood and sexual behavior, and oxytocin has been found to promote maternal bonding and affection for offspring, especially during nursing (Nelson & Panksepp, 1998).

One intriguing account of all this is that electrical brain activity in the form of potentials may underlie or contribute more to cognitive processes. Hormonal and synaptic action in the form of neurotransmitter function may then be the basis for or contribute more to emotion. In other words, thought may be *electrical*, while emotion may be *chemical*. If this were the case, then the same pathways that produce thought might also give rise to affect, because axonal and synaptic action are both necessary links in the process of neural information transmission. A particular pattern of neural activity in our brains could then simultaneously convey our experience of thought and affect. This corresponds with the earlier notion that the two are inseparable.

If mental computation—broadly speaking to include both cognition and affect—was expanded to include chemical and electrical activity, then the computing power of the brain may be significantly higher than previously estimated. Rather than 100 billion neurons with 1,000 connections each, we would need to add in a billion or so proteins in each neuron. This would yield a figure of a billion times 100 billion computing units (Thagard, 2008).

HOT AND COLD: EMOTION–COGNITION INTERACTIONS

Thoughts and emotions seem different. A thought seems "cool" and is associated with calm, measured deliberation. Emotions, on the other hand, seem "hot." They appear to drive behavior, impelling us to do something right now. However, this distinction may be a false one. A growing body of evidence indicates that emotions are part and parcel of many cognitive processes (Adolphs, Tranel, & Damasio, 1998; Barrett, 2017). It seems that emotion and cognition work together integrally in determining behavior. The neural mechanisms underlying both, for example, are connected at many different levels of

organization (Phelps, 2006). In the sections that follow, we wish to show that emotion–cognition interaction is complex. Following Thagard (2008), we reject the classical view that emotions always get in the way of thinking. We also reject the romantic view that emotions are better than reason. Instead, we adopt a critical view that shows how each can facilitate or interfere with the other.

Emotion and Perception/Attention

One of the proper functions of perception is to filter out information that is irrelevant and to focus on what is important. If a tiger is running at you, you certainly don't want to waste time looking at the flowers. Attention serves this function. Items that are allocated attention in the visual field are given priority and get processed more quickly, thereby enabling faster and perhaps lifesaving responses. But how does attention get allocated? There is evidence now to show that emotional processes can help steer attention toward significant stimuli—those that may be crucial for the organism's survival (Vuilleumier, 2005).

Using a visual search paradigm, Hansen and Hansen (1988) found that participants were faster at detecting an angry face hidden among neutral faces than they were at detecting a happy face under the same conditions (see Figure 10.3). These results have since been explained as an example of the **threat-superiority effect**, whereby dangerous items are found to be more perceptually salient. Easier detection of angry faces has since been replicated a number of times (Eastwood, Smilek, & Merikle, 2001; Fox et al., 2000). Notably, easier detection is not found for sad faces, indicating that it is the threat value rather than the negative valence that drives the effect (Öhman, Lundqvist, & Esteves, 2001).

Not only do threatening faces elicit greater perceptual saliency, but they also elicit greater perceived physical size. Holbrook and Fessler (2013) showed participants faces of recently successful and unsuccessful terrorist leaders and asked them to estimate their body size. The successful terrorist faces were estimated to have larger body sizes than the unsuccessful ones.

Whereas the aforementioned work shows a facilitative effect of emotionally charged stimuli, other work shows that it can interfere as well. Recall the Stroop effect, where it takes longer to *name the color of the ink* for a word that spells out a different color name (e.g., black ink spelling out the word *blue*) in contrast to that word printed in the same color of ink that it spells out. Pratto and John (1991) presented negative and positive adjective words in various colors. Their participants took more time to name the colors in the negative condition compared with the positive condition. In other words, it took more time to name the color of the ink for a word such as *ugly* than for a word such as *pretty*. The results suggest that our perceptual/attentional systems are oriented toward negative social information, and this can interfere with a primary task.

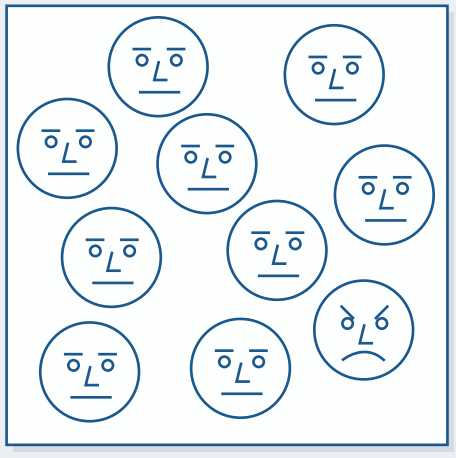
This slowing down in the Stroop effect implies that we are more sensitive to such information. We thus have a harder time ignoring it in the Stroop task. There is additional evidence supporting this notion of increased sensitivity. Phelps, Ling, and Carrasco (2006) presented either fearful or neutral faces and then followed these very quickly with four striped patches. The task was to judge which of the four patches was at a different orientation. Observers were faster at discriminating the orientation following the presentation of the fearful faces in comparison with the neutral ones.

There is also evidence that color can directly influence affect. Valdez and Mehrabian (1994) had people look at a color patch and then complete a questionnaire that measured their sense of Pleasure–Displeasure, Arousal–Nonarousal, and Dominance–Submissiveness. They found that different colors had different effects on people's affect. For example, blues, greens, and purples tend to elicit pleasant emotional reactions, whereas yellows tend to elicit unpleasant ones. Moreover, greens tend to elicit more general arousal, whereas purple and orange tend to elicit more relaxation. In fact, the results even suggested that red–purple induces a sense of submissiveness compared to green–yellow. These effects of colors on emotion are subtle but reliable across many participants (at least among the population of American participants on whom it was tested).

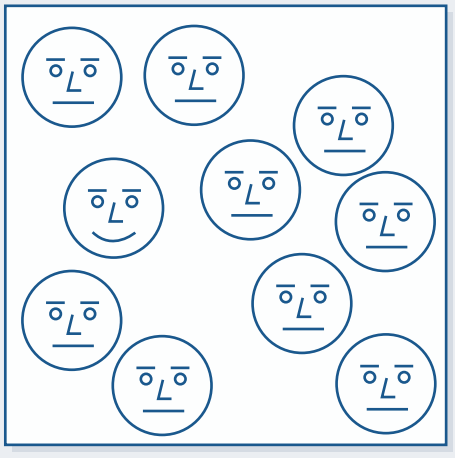
Emotion and Memory

What might be the role of emotion with regard to memory processes? One obvious prediction is that we should remember events surrounding emotional situations better. Since emotional events may be important to our survival, it makes sense that our memory systems might be set up this way. Imagine being chased by a tiger. If you climbed up a tree to escape and have a strong recollection of that event, then you would be more likely to employ the same strategy the next time you find yourself in that situation. So there is a clear evolutionary benefit to this type of memory system.

Figure 10.3 According to the threat-superiority effect, it is easier to detect an angry face hidden among neutral faces than it is to detect a happy face hidden among neutral faces.



Find the angry face.



Find the happy face.

Better recall of personal events during important or emergency situations is known as **flashbulb memory** (Brown & Kulik, 1977). People's recollections of what happened to them during the *Challenger* space shuttle disaster, the O. J. Simpson trial, and the attacks of September 11, 2001, are self-reported as being quite vivid (Reisberg, Heuer, McLean, & O'Shaughnessy, 1988). However, the accuracy of such memories is controversial. People have a high confidence in recalling what happened to them in these situations and report that they were very emotional at the time, but their memories may not actually be any better than during neutral times (Neisser & Harsch, 1992; Talarico & Rubin, 2003).

Flashbulb memories are for what a person was doing at the time of the incident. This includes where they were, whom they were with, and what they did. Such personal recollections are examples of what is called **autobiographical memory**. Whereas semantic memory is for facts, autobiographical memory is for episodes or events. This is thus considered a form of an episodic long-term memory, discussed in Chapter 5.

This means that we can now ask a question: Is memory for the emotional stimuli themselves, rather than the events surrounding them, improved? The answer appears to be yes. Ochsner (2000) presented participants with positive, negative, and neutral pictures and words. In a recognition paradigm, they were later asked to judge which of the stimuli they had seen before. If they thought they had seen an item before, they were then asked whether they remembered the context in which it occurred (the "known" condition, more episodic) or simply whether it was familiar (the "remembered condition," more semantic). The negative stimuli were remembered better than the positive or neutral stimuli. The arousing emotional words, positive and negative, were also remembered better than the neutral words. Furthermore, "remember" judgments were made more to negative stimuli.

There are several reasons why emotional items might be better remembered (Fox, 2008). First, it could be that we allocate more attention to such items. Things we pay more attention to are remembered better. Second, it could be because emotional things are more distinctive and unusual. In other words, they stand out more because they are rare and strange. A third alternate account is that we might think about them more often and thus rehearse them more. It may be that all three of these contribute to varying degrees, although at least one study has found better recall for negative affect items that cannot be entirely due to distinctiveness (Christianson & Loftus, 1991).

Emotion, Mood, and Memory

In the introduction, we distinguished between emotion and mood, stating that mood was less intense but longer lasting than emotion. A number of studies have examined the influence that mood has on memory (Levine & Pizarro, 2004). The research on this topic has focused on two effects: (1) mood-congruent memory and (2) mood-dependent memory. We will examine each of these next.

Mood-congruent memory is when we remember more stimuli if those stimuli match a mood we were in while learning those stimuli. Studies investigating this induce a particular mood at encoding while a participant is studying items, and then later, at recall, we see if the performance is better for matching items. The results support a

mood-congruency effect. People are generally better at recalling negative words such as *death* if they learned them in a negative mood and words such as *flower* if induced into a positive mood at encoding (Rinck, Glowalia, & Schneider, 1992). One explanation for this is that we think more deeply and make more associations to mood-congruent material (Bower, 1981).

In **mood-dependent memory**, recall is better when the mood at recall matches that during learning. Here, the same mood must be present at both occasions for there to be an effect. Note the difference here. In mood congruency, the stimuli match the mood at encoding. In mood dependency, the moods at encoding and at recall have to match. Bower, Monteiro, and Gilligan (1978) found that neutral words were better remembered when the mood at encoding matched the mood at recall. The two induced states were happy and sad. In these cases, it may be that mood is acting as a retrieval cue; the learned items became associated with the mood at encoding and reactivated this association at retrieval.

Fiedler (2001) argues that we engage in different information-processing strategies while in different moods. When we are in a good mood, we are more likely to assimilate. In **assimilation**, we take information and interpret it in terms of what we already know. This is a less critical strategy and stands the chance of being wrong. For example, we may incorrectly categorize a scooter as a motorcycle, failing to note the differences between them. On the other hand, when we **accommodate**, we change our internal concepts or schemas to take new information into account. In this example, we would form a new category for "scooter," differentiating it from "motorcycle."

According to Levine and Pizarro (2004), a positive mood corresponds to a motivational state where we have attained a goal and there is no immediate problem to be solved. Under such conditions, our cognitive processing is a bit "looser" and accuracy is not as important. A negative mood, in contrast, corresponds to a situation where we encounter a problem and a goal is threatened. Here, we need to "tighten up" our processing, becoming more analytic and detail oriented.

Emotion and Decision Making

Thagard (2008) outlines some of the differences between emotional (intuitive) and rational (calculated) decision making. Intuitive decisions are fast; they can be made immediately and lead directly to a decision. This is preferable when making less important judgments, such as whether you should order out for pizza or Chinese food. However, they are prone to error because they may be based on inaccurate or irrelevant information. When making calculated decisions, we usually consider more alternatives and goals, understand the pros and cons of each, and make the entire process explicit so that we can go back and rethink it or get somebody else to help. This process is slower but is the way we would want to approach important problems in our lives, such as whether or not we should marry someone or accept a job offer.

Damasio (1999) argues that there are situations where emotion facilitates decision making. Several of his patients suffered damage to the ventral and medial portions of the prefrontal cortex (vmPFC). Following this, they lost the ability to apply rationality to personal and social situations. He believes that the damage prevented them from

making emotional evaluations that involve somatic markers. These are bodily states that indicate whether a particular outcome will be good or bad. It is important to note that these patients could still reason well in other contexts and in some cases scored well on cognitive tests.

Bechara, Damasio, Tranel, and Damasio (1997) introduced a clever way to measure the role of emotions in decision making. In the Iowa Gambling Task, players received $2,000 in fake money to spend. They were presented with four decks of cards. In the first two decks, turning over a particular card resulted in a $100 win. However, many cards in these decks also produced large losses. In the other two decks, the wins were smaller. Turning over a winning card in those decks resulted in only a $50 gain, but there were no associated losses from turning over other cards. The best way to win and maximize earnings was to pick only from the latter two decks without associated losses.

Normal participants who played this game eventually realized (consciously or subconsciously) that the first two decks were risky, and they eventually stopped picking from them. In comparison, patients with bilateral damage to the vmPFC never stopped drawing from those decks, even when they were aware that they were risky options. This demonstrates that the vmPFC helps us determine whether a particular course of action is worth pursuing based on our past experiences. Normally, if we've done something and it resulted in a positive outcome, we will continue to do it. If another action has negative consequences, we will stop doing it. The emotional processing that enables us to do this is apparently missing in these patients, even if they were consciously aware of the risk.

Other research supports the notion that emotions are part of our normal decision-making process. Rather than believe that emotions are separate from cognition, it is more realistic to say that emotions are an integral part of them and help us choose from among alternatives (Loewenstein, Weber, Hsee, & Welch, 2001). Fazio (2001) reviews literature showing that most concepts have attached emotional attitudes. This certainly seems to be the case. It would be hard to imagine thinking about topics such as sex, politics, religion, and our friends and family without some sort of emotional bias. As we indicate in greater detail in the neuroscience section, the brain's emotional and cognitive centers work together during decision making (Rolls, 1999).

Clore, Gasper, and Garvin (2001) additionally advocate that feelings as well as information are used in decision making. For instance, if you find yourself in a good mood when in the presence of a certain someone, you may take that as evidence that you like him or her and decide to ask that person out on a date. This suggests a whole new class of process models, ones where feelings are combined with thoughts to determine our decisions and behavior (Fox, 2008).

Emotions and Reasoning by Analogy

Recall from the introductory chapter that analogies can be considered a type of mental representation. They can be used in problem solving and decision making by applying what we know in one situation to another similar to it. Thagard (2008) lists three types of analogies defined by their relation to emotion. First, there are analogies that are about emotions. Second, there are analogies that involve the transfer of emotion between

situations. Third, there are analogies that generate new emotions. We detail each of these types below.

We often use analogies to describe our mental states. Examples include being "as brave as a lion" or "sneakier than a rat." Perhaps this is because it is difficult to describe what a thought or feeling is like in literal terms. Analogies of the second type that are used to transfer emotions are found in persuasion, empathy, and reverse empathy (Thagard, 2008). We often use analogies when we are trying to persuade somebody. These sorts of analogies are quite prevalent in the news media (Blanchette & Dunbar, 2001). Examples include "living on easy street," rumors spreading "like wildfire," and being led "like pigs to a slaughter."

In empathetic analogies, we express our ability to appreciate somebody else's situation by comparing it with another similar situation. For example, we might say, "The Korean war was awful. The war in Vietnam was like it in several ways. So the Vietnam war must have been awful, too." In reverse empathy, we make this relation personal: "I was a victim of the New York City attack. When you were a victim of the Washington, D.C., attack, you felt terrified. So I am feeling terrified as well."

In the third case, analogies are used to generate new emotional states. According to Thagard (2008), these involve humor, irony, and motivation, to name just a few. Here is an example he gives to illustrate humor: "How can a single woman get a cockroach out of her kitchen? Ask him for a commitment." As a joke, this ought to generate feelings of surprise and amusement in the listener. Ironic analogies can be used to evoke negative emotions as well, as when we say that a failing bank's efforts are like "rearranging the deck chairs on the *Titanic*." We can also use analogies to motivate, as when a father compares his son with Albert Einstein or Mozart.

EMOTIONS AND ARTIFICIAL INTELLIGENCE: AFFECTIVE COMPUTING

Affective computing is computing that relates to, arises from, or deliberately influences emotions and other affective phenomena (Picard, 1997). The goal of such projects is to enable computers and robots with the ability to recognize emotions in people. This capability would facilitate human–machine interactions. For example, your computer could tell when you are tired or frustrated based on your facial expression. It might then recommend that you take a rest break. Secondary, but also interesting, is imbuing machines with the ability to express emotions. Emotionally expressive machines are more satisfying to interact with. The Kismet project is a first step in this direction (Breazeal, 2002). Kismet is capable of expressing a wide variety of facial expressions and can carry out extended, lifelike social interactions. See the "Interdisciplinary Crossroads" section for a detailed description of this project.

A number of computer programs already exist that are capable of recognizing human emotion from spoken auditory information. Petrushin (2000) describes the

emotion recognition project that uses an artificial neural network program to process features present in the speech stream. Examples of these features are energy, speaking rate, and the fundamental frequency. This program can recognize five emotions—(1) anger, (2) fear, (3) happiness, (4) sadness, and (5) a neutral, emotionless state—with about 70% accuracy. This is equivalent to human-level performance. There are plans to implement this system in automated telephone reservations systems such as those used by airlines to book flights.

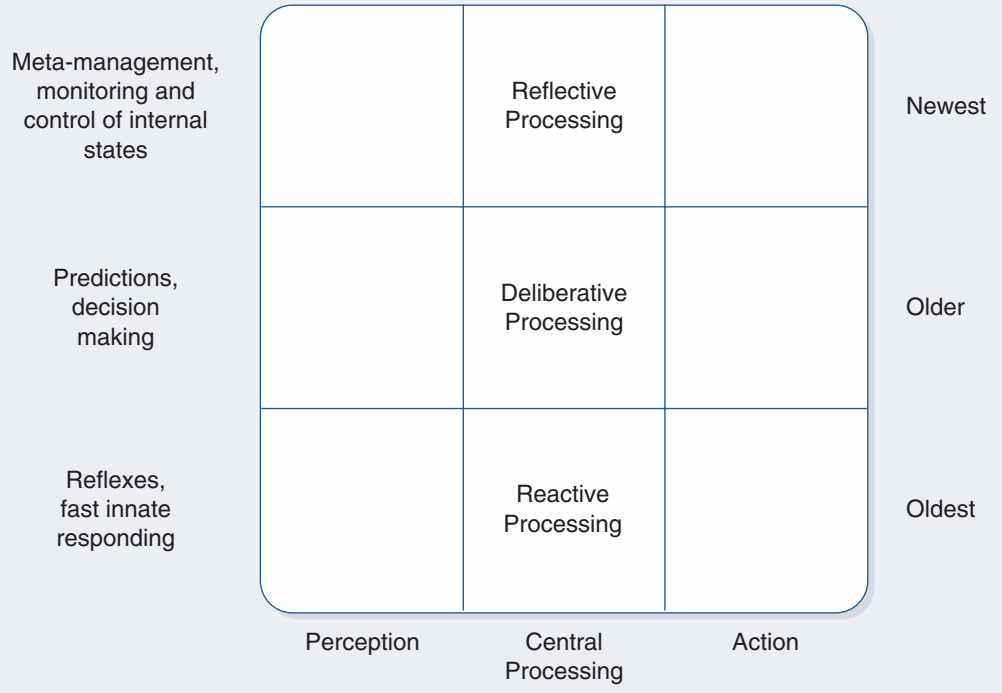
Researchers in the affective computing group at the Massachusetts Institute of Technology are developing a wide variety of affective systems. Their FaceSense program can tag facial expressions, head gestures, and affective–cognitive states from real-time videos for use in human–machine and human–human interactions. The system uses a combination of bottom-up visual cues, such as frowns, and top-down predictions of what expected affect should be like given the situation and context. It can be used to evaluate customer satisfaction and as a training tool for people with autism spectrum disorder. Another model this group is developing is called SmileSeeker. It also captures facial cues to help interpret social interactions. The current version specifically looks for smiles in face image data. This program could be implemented for use in bank teller training.

Recognizing and producing emotions is just one facet of affective computing. Other investigators are focusing on what the internal design of an emotional being is like. The **CogAff architecture** is a generic scheme for how cognitive–emotional processing may take place in either a person or a machine (Sloman, Chrisley, & Scheutz, 2005). In this model, there are three distinct levels of processing (shown in Figure 10.4). The first and the lowest level is **reactive**. These are reflex-like responses in which a stimulus automatically triggers a reaction—for example, the fear experienced after bumping into something in the dark.

The second stage involves **deliberative** processes. These involve representations of alternate plans of action. Here, different courses of behavior are generated given the circumstances, and the best one is selected. For instance, one could determine whether or not to go on a rollercoaster ride at an amusement park. Emotional states such as fear and cognitive factors such as one's medical condition are both taken into consideration.

In the third, **reflective** stage, we see metacognition come into play. Recall that metacognition is the ability to think about and control other mental states. At this level, one is aware of experiencing an emotion but can choose to act on it or not depending on the circumstances and the use of deliberative processes. An illustration of this would be deciding whether or not to give a speech at a conference. One could be afraid of doing so based on the possibility of being critiqued or evaluated. After reflecting on this fear, though, one might decide to go ahead and do it anyway because it would further his or her career. Alternate options, such as writing a paper, would be considered along with their associated emotions before a decision is reached. When giving the talk, one would then suppress fear but know why he or she is doing it.

Figure 10.4 The CogAff architecture specifies how a machine may exhibit emotional behavior.



Source: Adapted from Sloman et al. (2005).

INTERDISCIPLINARY CROSSROADS: EMOTION, ROBOTICS, AND THE KISMET PROJECT

Researchers in robotics are beginning to realize the importance of emotion (Breazeal, 2002). The goal of this new area in robotics is to design robots that both recognize and express emotions in much the same way that humans do. Robots with these capacities will make it easier for us to interact with them. The **Kismet** project is the first major step in this direction. It is designed to model the interaction between an infant and its caregiver. Kismet is a cute robotic head capable of sensing others and of expressing a wide range of facial expressions. It is driven by a cognitive

and emotional system that work together to govern its interactions with people (Breazeal & Brooks, 2005).

Kismet can "see" using a set of color cameras. It can move its head and eyes to control where it looks and what it pays attention to. Kismet's auditory system is a microphone that can process and recognize certain aspects of human speech. The auditory system allows it to detect basic pitch contours that signal approval, prohibition, attention, and comfort (Fernald, 1989). The detection of this affective information then guides its own emotional state. Kismet can fold back its ears to signal anger or raise them to display a state of interestedness. It can also move its eyebrows up and down or furrow them to communicate surprise, frustration, or sadness. In addition, Kismet is equipped with a vocalization system, allowing it to generate synthesized sounds reminiscent of a young child.

The Kismet robot conveys emotions mostly through its face (see Figure 10.5). The emotions it displays have been formulated to fit within a three-dimensional affect space of valence (good or bad), arousal (high or low), and stance (advance or withdraw). A soothed expression corresponds to high positive valence and low arousal (i.e., a state of being happy but not excited). A joyous expression instead corresponds to a state of positive valence and moderate arousal. Some of the other expressions Kismet is capable of include anger, disgust, sorrow, surprise, and fear.

Kismet's cognitive system consists of perception, attention, drive, and behavior subsystems. It is motivated by basic drives just as a real biological child would be. These drives are thirst, hunger, and fatigue. A social drive gives it a "need" to interact with people, a stimulation drive provides an impetus to play with toys, and a fatigue drive causes it to rest when it is overstimulated. When these "needs" are met, the drives are in a homeostatic regime and Kismet acts as though it is satisfied. But if the intensity level of a drive deviates from this state of balance, the robot is motivated to engage in behaviors that restore the drive to equilibrium. Kismet's drives don't directly produce an emotional response, but they do, however, bias its overall emotional state or mood.

If Kismet does not receive enough stimulation and its social drive state is high, it puts on a sorrowful expression. Under typical circumstances, this will elicit increased interaction by a human playmate. For example, a person might try to engage Kismet by waving a toy in front of its face. If the robot is getting too much stimulation, its drive state is lowered, causing it to make a fearful face. In this case, a normal human would back off to reduce the amount of stimulation. If Kismet is receiving a moderate amount of stimulation, it expresses joy to encourage sustained human interaction.

Kismet can use its eyes and vocalize to communicate as well. If it is presented with an undesired stimulus, a disgust response will reorient its gaze to an alternate area in the visual field, where it might locate a more desirable object. This would again serve as a cue to a person to change his or her behavior, perhaps by switching to a different toy. If a person is ignoring Kismet, it can attract the person by first vocalizing. If that doesn't work, it might lean forward and wiggle its ears to attract attention. So, just like a real child, Kismet can use a variety of behaviors to get what it "wants."

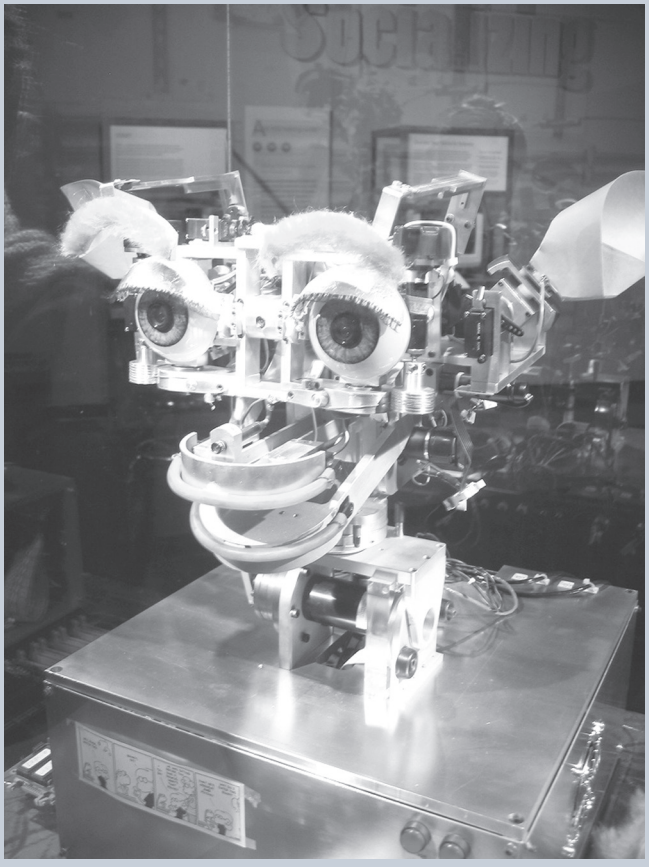
Kismet's creators programmed it with several basic-level emotional states, including interest, surprise, sorrow, joy, calm, fear, disgust, and anger. The trigger state for one of these particular emotions corresponds to the values of the different motivational drives. The activation of the emotion would then, in turn, trigger the facial expression that goes along with it. A human observer, seeing this display, would then

(Continued)

(Continued)

modify his or her actions toward it. People enjoy playing with Kismet and very easily attribute humanlike qualities to it. Future robots with emotionally perceptive and expressive abilities like Kismet will undoubtedly make it easier for us to accept and work with them.

Figure 10.5 Kismet the robot is designed to interact with people in humanlike ways.



Source: Jared C. Benedict, CC-BY-SA.

OVERALL EVALUATION OF THE EMOTIONAL APPROACH

The emotional approach is a recent but welcome addition to cognitive science. For too long, researchers have swept emotions "under the rug" and focused exclusively on cognitive factors. It is now generally accepted, though, that most cognitive processes work in concert with affective ones. The research we have reported here is a good start. We are just beginning to understand how emotions affect perception, attention, memory, language, and problem solving. Many questions remain, however. We need a more detailed description of the physiology and information flow that underlies each of the emotions and moods.

Most of the research reported here looks at the influence of emotion on cognition, presumably because these cognitive processes have been studied for some time and are better understood. But the reverse situation, the influence of cognition on emotion, deserves research attention as well. For instance, how do we suppress emotions, and under what conditions is this easier or more difficult? There seem to be large individual differences in this ability, with some people predisposed to be more emotional and others more rational. If we understood the neural mechanisms underlying aggressive behavior, we might be able to reduce theft, murder, spousal abuse, and other crimes.

Understanding the basic emotions is just the tip of the iceberg. Many other more subtle and complex emotions remain. If these are created by a combination of the basics, then are there lawful rules governing this process? Perhaps anger and admiration mix to produce envy in much the same way that blue and yellow combine to make green. These "secondary" emotions play a powerful role in our thinking and behavior, but they have yet to be well classified or explained. There may even be "tertiary" emotions that are combinations of affect and cognition. Arrogance and self-esteem seem to fall into this category.

SUMMING UP: A REVIEW OF CHAPTER 10

1. An emotion is a relatively brief brain and body episode that facilitates a response to a significant event. Feelings correspond to the subjective experience of an emotion. Moods are less intense but longer lasting types of affective states.
2. Theories of emotion all posit a stimulus that triggers physiological changes and some sort of appraisal. However, these theories disagree on the temporal ordering of these events and the exact role that the cognitive appraisal process takes.
3. Some of the basic emotions that may be universal across cultures are happiness, sadness, fear, anger, surprise, and disgust.

4. Each of the major emotions may have played an adaptive role in promoting our survival during the conditions of our ancestral past. However, problems with these emotions can also be linked to psychological disorders.
5. The amygdala is a brain structure that mediates our fear response. It also plays a role in improving our memory for arousing stimuli.
6. More contemporary work in affective science has investigated the role that emotions play in perception, attention, memory, and decision making.
7. Emotions focus our attention on dangerous events. This is known as the threat-superiority effect. They also bias us to remember events of emotional significance. Under certain conditions, emotions can even facilitate decision making.
8. A number of computer models already exist that are capable of recognizing emotions based on cues in speech or facial images.
9. The CogAff architecture is an example of how an organism or machine may integrate emotions into decision making. It uses reactive, deliberative, and reflective stages.

SUGGESTED READINGS

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Boston, MA: Houghton Mifflin Harcourt.

Fellous, J. M., & Arbib, M. A. (Eds.). (2005). *Who needs emotions? The brain meets the robot*. Oxford, England: Oxford University Press.

Fox, E. (2008). *Emotion science*. Basingstoke, England: Palgrave Macmillan.

Lane, R. D., & Nadel, L. (2002). *Cognitive neuroscience of emotion*. New York, NY: Oxford University Press.

Levy, D. (2007). *Love and sex with robots: The evolution of human-robot relationships*. New York, NY: Harper.

Minsky, M. (2006). *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. New York, NY: Simon & Schuster.

Solomon, R. C. (Ed.). (2003). *What is an emotion? Classic and contemporary readings*. Oxford, England: Oxford University Press.

Thagard, P. (2008). *Hot thought: Mechanisms and applications of emotional cognition*. Cambridge: MIT Press.

CHAPTER ELEVEN

THE SOCIAL APPROACH

Mind as Society

SOCIAL COGNITION

For much of the early history of cognitive science, the focus was on the individual. Cognition was seen as something that was generic—whether one thought about a tractor, a tree, or one's best friend didn't make a difference because the cognitive processes were assumed to be the same. We now know that this is not necessarily the case. There is evidence to show that we think about people in ways that differ from the ways in which we think about inanimate objects. This makes sense. We are social animals and have evolved in social settings where we interact with others. To quote John Dunne, "No man is an island unto himself, he is a piece of the main." In this chapter, we will examine how we think about people from an interdisciplinary perspective, examining the role of evolution, the underlying neural mechanisms, and the cognitive models.

Social cognition is the study of how people make sense of other people and of themselves. Fiske and Taylor (2008) state that there are four assumptions that pervade the field. The first is mentalism, which is that people use cognitive *representations*. This is nothing new, except that in this case, the things that are being represented and computed on are social in nature. They include people's faces and appearance, as well as the feelings, thoughts, and explanations we have about them. As we shall see, we often think differently about people than we do about inanimate objects, such as chairs, or abstract concepts, such as justice. Table 11.1 summarizes some of the important differences in thinking about people versus things.

The second basic assumption in research on social cognition is about cognitive *processes,* or how representations develop, operate, and change over time. So if we were studying prejudice, we would want to know not just what prejudice is like but also how it forms

Learning Objectives

After reading this chapter, you will be able to:

1. List the four vertices of the prism model of social cognitive neuroscience.
2. Contrast the cultural intelligence hypothesis with the general intelligence hypothesis.
3. Discuss the social role of joint attention during language learning with children.
4. Describe what the mirror neuron system is.
5. Explain what theory of mind refers to.
6. Describe the symptoms associated with autism spectrum disorder.
7. Explain what cognitive dissonance is.
8. Describe the evidence for automatic stereotyping.

and how it might be reduced or prevented. The third theme is cross-fertilization, which is that the social cognitive approach is highly interdisciplinary in nature. In fact, most of the work we will report in this chapter occurs at the intersection of the cognitive, social, and neuroscience perspectives. We will discuss the intertwining of these three perspectives in more depth later.

The fourth theme is the applicability of social cognitive research to the real world. Since we live in a social society, we can use the findings we obtain to make the world a better place. For example, our understanding of stereotypes may enable us to reduce the development of negative stereotypes and enhance the formation of positive ones. Results from social cognitive research have been applied to many areas, including health care, the legal system, advertising, and political campaigns.

There is a lot to say about social cognition. To organize this material and make it easier to digest, we will divide the chapter into three sections. In the first section, we will talk about social cognitive neuroscience. In the second section, we will examine

Table 11.1 Differences in the way we think about people and things.

Principle	People (Example: Your Best Friend)	Things (Example: A Rock)
Intentionality	People attempt to control the environment.	Objects are not capable of controlling their environment.
Mutuality	People think about each other (e.g., "I think of you, and you think of me").	Objects don't "think back."
Self-concept	Others judge us, provide information about us, and are more similar to us than are objects.	Objects cannot judge or provide information about us and are very different from us.
Observational bias	People act differently when they know they are being thought about.	Objects do not act differently when they are being thought about.
Nonobservable attributes	Traits that can't be observed are crucial to thinking about people.	Traits are more easily observed but less crucial.
Change	People change over time and in different circumstances.	Objects change less over time and in different situations.
Accuracy of cognition	It is difficult to judge the traits or qualities of a person.	It is usually easier to judge object qualities.
Complexity	People are complex; we need to simplify in order to understand them.	Objects are usually less complex.
Explanation	We have to explain others' behaviors.	Explanations are not always necessary.

Source: Adapted from Fiske and Taylor (2008).

perhaps the most important foundational aspect of social cognition. This is the theory of mind (ToM), which is how we come to know that other people have minds. In this section, we will discuss the role ToM plays in social cognitive disorders such as autism. Finally, we will sum up by covering some of the basic social topics that explain how we think about others. These topics are impressions, attributions, attitudes, stereotypes, and prejudice.

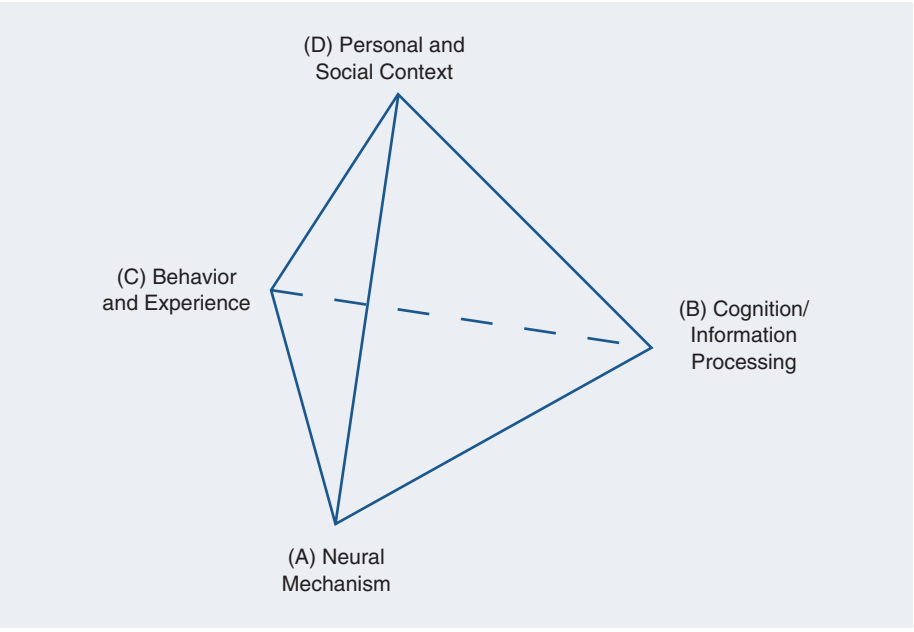
SOCIAL COGNITIVE NEUROSCIENCE

A common theme throughout this book is the interdisciplinary approach, where different perspectives and methodologies focus on a research topic. No area illustrates this more clearly than the field of social cognitive neuroscience. As the name implies, this endeavor attempts to describe phenomena in terms of interactions between three levels of analysis. At the broadest level, we have social psychology, which examines the role of human interaction, focusing on people's behavior in a group setting. At the intermediate level, there is the cognitive perspective, examining how we think about one another. Ultimately, we arrive at the neuroscience level, which seeks to examine the location and type of neural processing that takes place in a social setting and that underpins and gives rise to our cognition and behavior.

Ochsner and Lieberman (2001) make several important points about this young field. They note that emphasis needs to be placed on the cognitive approach, because it is here where social psychologists and neuroscientists have the most in common and can most easily communicate. Cognition serves to "bridge the gap" and unites the fields of social cognition and cognitive neuroscience. They also note that each of these areas can enrich the other, with social psychologists using neuroscience findings to clarify and differentiate between competing psychological theories. On the other side of the divide, neuroscientists can use findings from the social literature to guide their exploration into whether there are different neural mechanisms that underlie social versus nonsocial stimuli.

To illustrate the relationship between these three levels, Ochsner and Lieberman (2001) present a prism model (shown in Figure 11.1). Each corner or vertex of the prism represents one of four different disciplinary aspects. Each face is bordered by three corners and stands for one of the three approaches. For instance, the base of the prism is formed by neural mechanisms, cognition/information-processing, and behavior and experience vertices. This surface, which faces downward in Figure 11.1, corresponds to the domain of inquiry for cognitive neuroscience but is lacking because it does not take into account the remaining vertex, personal and social context. Similarly, social psychology is the facet formed from the personal and social context, behavior and experience, and cognition/information-processing vertices. This is the surface that faces toward the back in Figure 11.1. It, too, is lacking because it fails to account for the remaining vertex, neural mechanisms. Only the field of social cognitive neuroscience takes all these aspects into account and is thus represented by the entire prism with all the four of its faces.

Figure 11.1 The prism model of social cognitive neuroscience.



Source: Ochsner and Lieberman (2001).

TOPICS IN SOCIAL COGNITIVE NEUROSCIENCE

Evolution

We humans are a distinctly social race, what might be called "ultrasocial" (Boyd & Richerson, 1996). More than any other species, we form cooperative groups, such as family, friendships, and teams, as well as competing groups, such as tribes, nations, and ethnicities. Is it possible that this sort of social environment led to the biological and cultural evolution of our intelligence? According to the **cultural intelligence hypothesis**, it has. In this view, complex social interaction created a form of social intelligence that then laid the groundwork for more general intelligence to emerge later (Herrmann, Call, Hernandez-Lloreda, Hare, & Tomasello, 2009). For example, our working memory capacity may have evolved originally to allow primates to remember key members of their social group. Once this capacity was inherited and enhanced in humans, it could be applied to more abstract problems involving language, mathematics, and spatial reasoning.

In contrast, the **general intelligence hypothesis** states that human intelligence did not evolve in response to particular selection forces such as social organization. Instead, it emerged in a general form that could be suited to solving all kinds of problems, whether

social or not. One way to differentiate between these two notions is to examine the physical and social cognitive abilities of children who have not yet been influenced by cultural factors such as written language, symbolic mathematics, and formal education. Physical cognition refers to reasoning about things such as space, quantities, and causality. This could evolve in a wide variety of environments and is not specific to social settings. Social cognition, on the other hand, is reasoning about those things that are directly linked to culture, such as social learning, communication, and ToM.

We can then compare the children's performance with that of our closest primate relatives. If the cultural intelligence hypothesis is correct, there should be no difference between the children and the primates on physical domain tests because both groups have evolved under these conditions. There should, however, be a difference on tests of social cognition. We would expect the children to score higher on these tests because humans evolved in more demanding social environments.

This is exactly what Herrmann et al. (2009) did. They tested chimpanzees, orangutans, and 2½-year-old human children. The results clearly supported the cultural hypothesis. There was no difference between human and primate performance in the physical domain. On tests designed to specifically measure social intelligence, the children's scores were higher by a statistically significant amount. If humans had evolved a superior general form of intelligence, they would have outscored the primates in both categories. It is not known yet what specific social environments may have led to this divergence in our early ancestors, but it could have been complex types of collaborative activity involved in hunting and gathering that would call on more sophisticated kinds of communicating and social learning (Tomasello, Carpenter, Call, Behne, & Moll, 2005).

Additional evidence to support the evolution of specialized social intelligence comes from studies that have examined brain size. There is a positive correlation between the relative size of the neocortex and social factors such as group size and grooming clique size that are measures of social complexity (Dunbar, 1996). Hedgehogs are mammals that provide maternal care to a small number of offspring, and they have comparatively small cortical regions. Chimpanzees live in extended groups of a few dozen animals and show moderate cortical size. Humans, who interact in societies numbering in the millions, have the largest relative cortices. In studies such as this, it is not absolute brain weight or volume that is measured but, rather, relative size expressed as a proportion of the animal's overall body size. It could be that larger social groups cause greater competition for social skills that subsequently leads to the evolution of cognitive mechanisms for outsmarting others.

Attention

It has been known for some time that infants can follow somebody else's gaze (Scaife & Bruner, 1975). This suggests that even at a young age, children can focus their attention on what another person is attending. This ability to coordinate attention with a social partner is called **joint attention**. Not only following but also initiating or joining shared attentional focus with others is a key part of social interaction. Without it, we would have a very difficult time understanding or making sense of what other people are saying (Mundy & Newell, 2007).

We can divide joint attention into two categories. In responding to joint attention (RJA), we follow the gestures and direction of gaze in order to share a common point of reference. This is the more passive form, because we are following others. In initiating joint attention (IJA), people use gestures and eye movements to direct attention to people, objects, and events. This is a more active form, because it involves the initiation of a joint attentional focus.

Joint attention plays a very important role in development. It can help disambiguate the referent of a word during language acquisition. A mother pointing to a *spoon* and saying the word gets a child to look at the object corresponding to the word and so aids in the growth of vocabulary (Baldwin, 1995). In fact, even when the mother doesn't point to the object, research shows that when a child is holding the object in their hands, it dominates their field of view so much that there is very little uncertainty as to what the child will be looking at while they hear the name for that object (Smith, Yu, & Pereira, 2011). Unfortunately, children with autism show impairments primarily in IJA rather than in RJA. This may account for some of the developmental delays experienced by autistic children. We will cover autism in greater detail elsewhere in this chapter. Interestingly, chimpanzees have the capacity for RJA but not for IJA (Tomasello & Carpenter, 2005).

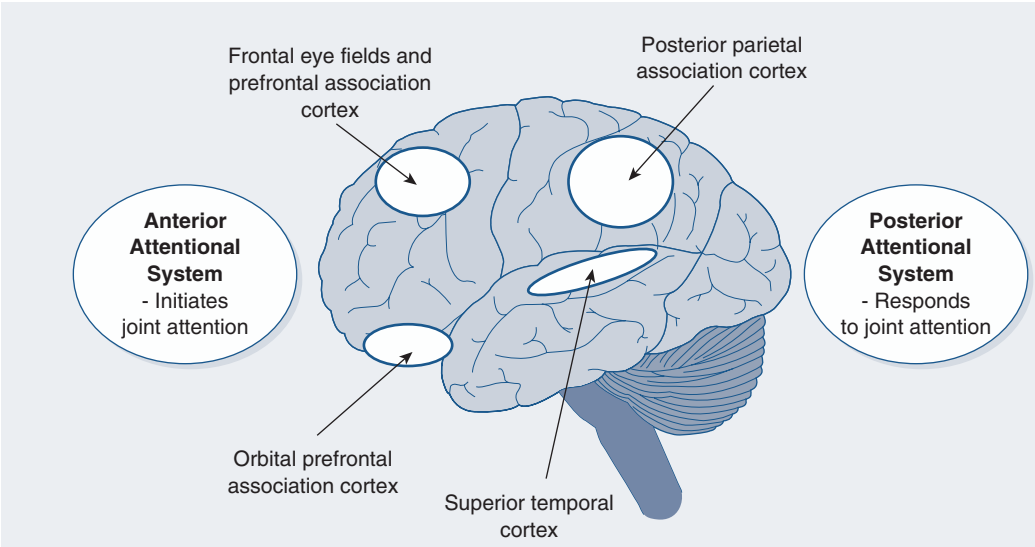
Social cognition begins around the age of 9 to 12 months. At about this time, infants should begin to be aware of their own goals and activities as well as those of others (Tomasello et al., 2005). Social cognitive skills allow the child to combine these. They thus realize that their intentions lead to actions and that this is the case for others as well. If I want to eat, then I look at or point to the cereal box because I want to eat it or get mom to help me eat it. If mom also points or looks at the cereal, then she, too, must be thinking about it and plans to prepare it for me to eat. In this fashion, children learn that thought is representational and directed toward objects and actions in the world, not just in themselves but in others as well. This realization can then lead to the use of language and communication to convey the contents of one's mind to others.

There are two attention regulation systems underlying joint attention (Mundy, Card, & Fox, 2000). These are depicted in Figure 11.2. The posterior orienting and perceptual system is used during the learning of RJA in infancy. It consists of the parietal and superior temporal cortex. This system is involuntary and is used to direct attention to biologically meaningful stimuli. This underlies our capacity to perceive the eye and head orientation of others as well as the spatial representation of our own bodies in relation to other people's bodies and objects in the environment. It allows us to pay attention to where people are looking or pointing to and so is a more basic or perceptual level of joint attention.

Second, there is the anterior attention system. This is subserved by a collection of different neural areas, including the frontal eye fields, which control eye movement; the orbital frontal cortex; the prefrontal association cortex; and the anterior cingulate. This system underlies our capacity for IJA and emerges slightly later in development. It controls voluntary, goal-directed attention and is what enables us to understand that our personal behavior is directed toward where we attend. This may be thought of as a more sophisticated conceptual level of joint attention.

Figure 11.3 shows the attention systems model of how these two systems interact to determine joint attention. Early in development, at the age of 3 to 6 months, the anterior and posterior systems interact with each other and with additional processes such as

Figure 11.2 The posterior and anterior attention systems underlying joint attention (after Mundy et al., 2000).



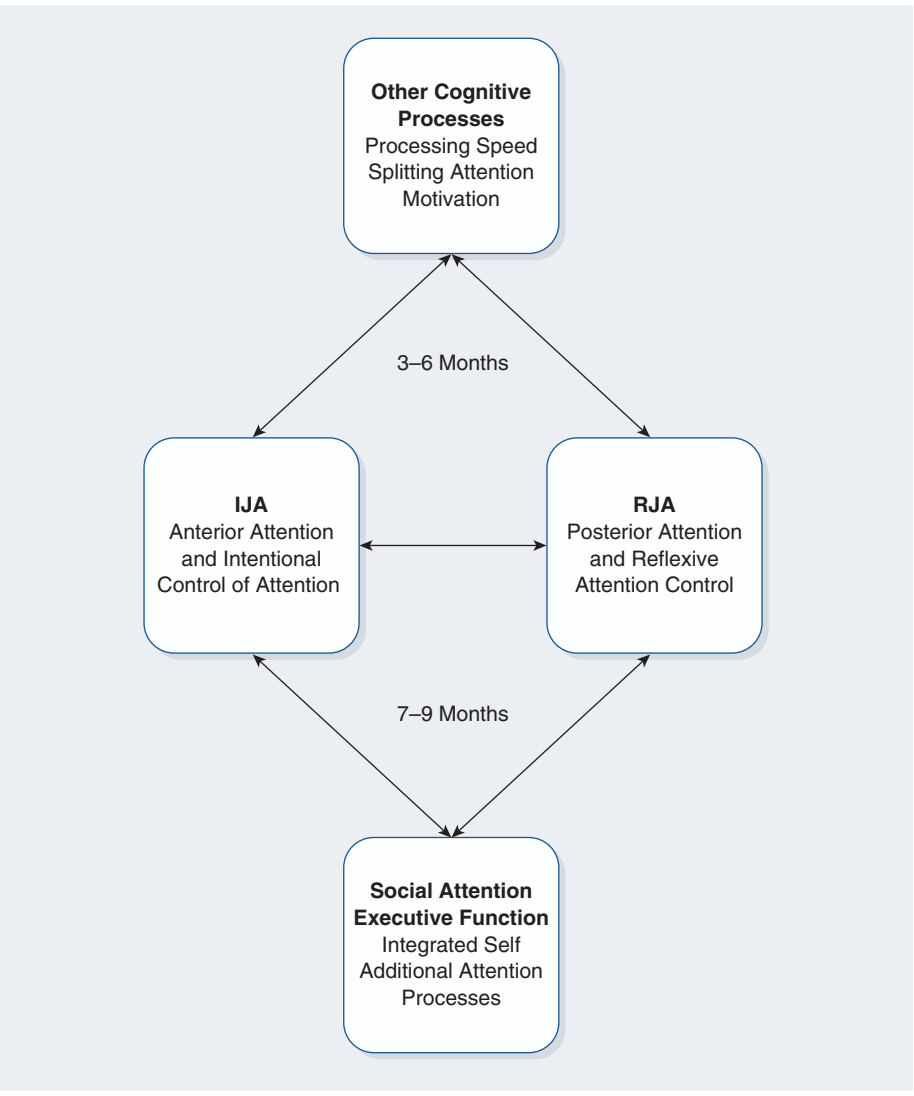
speed of processing, motivation, and the ability to divide attention. This dynamic leads to the development of integrated self–other attention processing at 7 to 9 months, which in turn serves as the foundation for the emergence of social cognitive joint attention at 10 to 18 months.

The attention systems model can account for two important observations mentioned previously. Cherkassky, Kana, Keller, and Just (2006) propose that problems in communication between brain areas, especially between the anterior and parietal areas, may explain autistic people's lack of social cognitive skills. In chimpanzees, the parietal attention system is represented, but the anterior system is not. This may explain why their RJA ability is better than their IJA ability (Astafiev et al., 2003).

Mirror Neurons

A **mirror neuron** is a neuron that fires when an animal performs some action and also when the animal observes another animal performing that same action (Rizzolatti & Craighero, 2004). So, for example, if a monkey were to reach out and grasp an object, we might see activity in particular neurons. These neurons would also fire when that monkey watches another monkey reach out to grasp the same object. In this sense, the neurons are said to "mirror" the performance of others. These cells could play a role in imitative or modeling behavior and help explain how some animals copy others as part of the learning process. A similar mirror neuron system has been observed in human

Figure 11.3 How the two attention systems interact to determine joint attention.



Source: Adapted from Mundy et al. (2000).

brains (Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, 2006; Stevens, Fonlupt, Shiffrar, & Decety, 2000).

In macaque monkeys, mirror neurons have been found in the inferior frontal gyrus and in the inferior parietal lobe (Rizzolatti & Craighero, 2004). In one study, 41 neurons were studied in two macaque monkeys while they observed an experimenter either grasp

an apple and bring it to his mouth or place it in a cup (Fogassi et al., 2005). They discovered 15 neurons that fired to the "grasp-to-eat" motion but did not fire to the "grasp-to-place" condition. Conversely, there were 4 other neurons that showed increased activity to the "grasp-to-place" motion but not to the "grasp-to-eat" motion. The cells fired in some cases before the final grasp was performed (i.e., before the apple was brought to the mouth or placed in the cup). These researchers speculate that these neurons, therefore, code for the animal's understanding of another's intention. That is, they become active when the monkey anticipates what somebody will do.

Several studies have examined this mirror neuron system in humans. Iacoboni et al. (1999), using functional magnetic resonance imaging (fMRI), found activity in the inferior frontal cortex and the superior parietal lobe when a person performs an action and also when the person sees another person perform that same action. The specific pattern of brain activity in these areas also has been found to be similar during the perception of others' actions and when performing the act oneself (Dinstein, Gardner, Jazayeri, & Heeger, 2008). Sebanz, Bekkering, and Knoblich (2006) have suggested that this mirror neuron system may play a role in helping people coordinate during joint action, even allowing them to have shared mental representations.

In addition to being used to model others' actions and behavioral intent, the mirror neurons system may also be used as part of the experience of empathy (Decety & Jackson, 2004). There is some support for this in imaging studies with human participants. Sections of the anterior cingulate cortex that are thought to mediate the subjective experience of pain are also active in individuals when they watch a video of somebody else about to experience bodily harm—for instance, a person who is about to cut his or her finger with a kitchen knife (Singer et al., 2004). Smelling a foul odor was also found to produce activity in the anterior insula, both when experiencing this directly and when watching somebody else express disgust to a smell (Wicker et al., 2003). However, these brain areas are not the same as those that mirror hand actions, and mirror neurons coding for empathy and emotional states have not yet been found in monkeys (see Table 11.2).

Social Cognition as the Brain's Default State

Mitchell (2008) proposes a speculative but fascinating idea, which is that social cognition may be so special that it is the default state for the brain. The areas implicated in social thinking, such as the medial prefrontal cortex (mPFC), the right temporoparietal junction (rTPJ), and the precuneus/posterior cingulate cortex, have an unusually high metabolic rate; they tend to burn energy faster than other brain areas when at rest (Gusnard & Raichle, 2001). A higher resting rate means that these areas remain "on" and ready for action even though we may not be recruiting them to perform some task. This suggests that we remain in "people mode" whenever we are not actively thinking about anything in particular.

Social brain areas have also been shown to deactivate when people are thinking about nonsocial topics. So if you are pondering the answer to an exam question that requires semantic and conceptual knowledge unrelated to people, you literally need to "turn off" the social parts of your brain, perhaps, so they don't interfere. This seems to be unique to social cognition. When thinking about a nonlinguistic problem, for example, we don't deactivate language areas. These regions simply remain at their lower resting

Table 11.2 Brain areas considered part of the mirror neuron system.

Brain Area(s)	Cognitive Function/Ability	Animal
Inferior frontal gyrus and inferior parietal lobe	Imitating and understanding other's behavior	Macaque monkey
Inferior frontal cortex and superior parietal lobe	Imitation and determination of behavioral intent	Human
Anterior cingulate cortex	Empathetic experience of pain	Human
Anterior insula	Empathetic experience of disgust	Human

metabolic rate. Mitchell (2008) concludes that we are in a state of constant readiness to encounter other minds, and this normal state must be shut down in order to reason about things that don't involve people.

IS SOCIAL COGNITIVE NEUROSCIENCE SPECIAL?

Social cognitive neuroscience is not without its critics. Some researchers question whether it has been able to provide new insights into the nature of cognitive processes (Willingham & Dunn, 2003). They question whether neuroimaging has yielded any novel information that could not be gleaned with more traditional methods such as accuracy and reaction time. Another question that has been raised is whether the general cognitive processes used in perception, attention, memory, language, and problem solving can explain social cognition or whether there are special cognitive mechanisms or neural processes that are specific to social interaction (Blakemore, Winston, & Frith, 2004).

Mitchell (2008) strongly defends the field against these accusations. He points out that a small number of brain regions active during social cognition are not active during most other nonsocial cognitive tasks. For instance, the mPFC is activated when participants are asked to read stories that require mental attribution, in comparison with those that involve physical causality (Fletcher et al., 1995). Numerous other studies also have found differential mPFC activation in social cognition tasks under circumstances where the stimuli, type of mental state, and behavioral tasks vary (Blakemore et al., 2004).

Additional evidence supports this rebuttal. Mitchell, Macrae, and Banaji (2004) performed fMRI scans on individuals while they read descriptions of people, and then, they were instructed either to form an impression based on the passage or to memorize the information. The impression formation task produced mPFC activity, while the memorization task did not. In addition, the mPFC activity was correlated with the likelihood of remembering material, but only when it was associated with the impression formation task.

However, research also shows that there may be some functional overlap of brain areas underlying social cognition and other more general cognitive operations. Saxe and Kanwisher (2003) found rTPJ activation when participants processed stories about another person's erroneous beliefs. This activation was not seen for thinking about stories that referred to an erroneous physical representation. But to complicate things, it turns out that the rTPJ is activated during tasks that require shifts of attention (Mitchell, 2008). So it turns out that this debate revolves around the specific type of social cognition performed. Unique brain areas are called on when doing some types of social thinking, but other forms of social mentalizing can recruit areas used in general cognitive processes.

ADVANTAGES OF THE SOCIAL COGNITIVE NEUROSCIENCE APPROACH

What are some of the potential benefits of this tripartite interdisciplinary venture? Ochsner and Lieberman (2001) advance three possibilities. They argue that social cognitive neuroscience could unify the field by discovering common brain mechanisms underlying what were previously thought to be disparate phenomena. If the same parts of the brain are used for attitudes and attributions, for example, then we may have to reconsider these not as two different processes but perhaps as two aspects of the same process. The amygdala, to give another example, is implicated in many social cognitive processes, including attitudes, stereotyping, emotion, and person perception. An understanding of what the amygdala does in each of these situations can unify these formally distinct research topics at least to the extent that they call on similar neural systems.

Social cognitive neuroscience may also have the opposite effect. In some cases, it could separate what might have been thought to be a unitary phenomenon. Currently, stereotypes are broadly categorized as implicit or explicit. However, it is likely that several different subtypes could include affective, semantic, declarative, and procedural stereotypes. It remains to be seen whether different neural mechanisms underlie each of these possibilities or whether there is a single or small number of distinct processes they share.

A third potential benefit of social cognitive neuroscience concerns bottom-up versus top-down approaches. By bottom-up approaches, we mean investigating the basic neural systems that service the perceptual, attentional, emotional, and behavioral processes of social experience. Top-down approaches refer instead to the diverse types of social phenomena, such as attitudes, attributions, stereotypes, and prejudice. It is generally agreed that more can be learned by integrating these two approaches than by treating them separately, as has been done most often in the past.

THEORY OF MIND

One of our most important social abilities, one that underlies all others, is the understanding that other people have minds. When you speak to your roommate, you assume that he or she is conscious and has feelings and thoughts in the same way that you do.

You don't treat him or her as though he or she is an inanimate object like a chair or table. This capacity is referred to as **theory of mind (ToM)**. It is the ability to understand others' mental states, such as beliefs, desires, and intentions; to appreciate how these differ from our own; and to use this understanding to guide our behavior when interacting with others.

A number of tests have been developed to diagnose ToM (Wimmer & Perner, 1983). The use of these tests shows that children acquire a ToM by the age of 4 to 5 years. In the Sally–Anne task, a cartoonlike drawing depicting two children is shown (Gallagher, 2000). Sally puts her marble in a basket. When she leaves, Anne takes the marble out of the basket and puts it into her box (Figure 11.4). Sally, now, returns. The question is, where will Sally look for her toy? Children without disorders, who are 4 years and older, will answer that he or she will first look for it in the basket. This requires an understanding of the situation from Sally's point of view—namely, that she holds the false belief that the toy is in the basket because that is where she last left it. A child lacking ToM would predict that she would look for it in the box, even though she could not actually know that's where the marble is because she was absent when Anne moved it.

ToM and Neuroscience

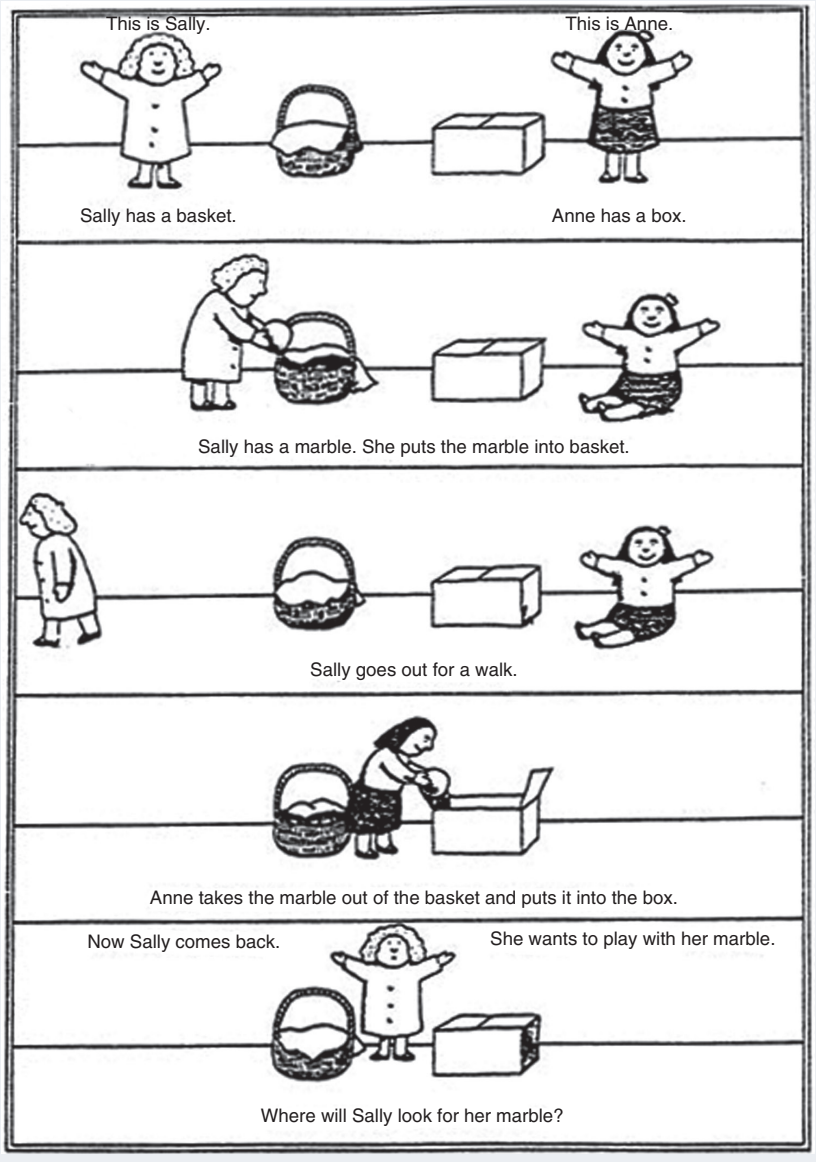
Siegal and Varley (2002) outline four neural systems that mediate ToM. These are (1) the language system, specifically the left hemisphere grammatical ability; (2) the frontal lobes; (3) the right temporal–parietal areas; and (4) the amygdala circuits. In this section, we will discuss the role each of these plays in ToM and address some of the issues concerning how ToM may develop.

An understanding of linguistic syntax may allow us to entertain false beliefs. De Villiers and de Villiers (2000) argue that propositional reasoning enables us to represent a person's beliefs and then modify these representations to express a new state of mind. For instance, in the Sally–Anne task, one could think that "Sally believed (falsely) that the marble was in the basket." Brain imaging studies show activation of language centers during the execution of ToM tasks. They found widespread temporal lobe activity, either bilaterally or localized to the left hemisphere language centers (Fletcher et al., 1995).

Frontal lobe executive function is also believed to play a prominent role in ToM reasoning. In many of the false-belief studies, suppression of the object's real location must occur to attribute the perceived location to one of the actors. Inhibition of inappropriate responses is one of the functions carried out by frontal areas (Carlson & Moses, 2001). Brunet, Sarfati, Hardy-Bayle, and Decety (2000) obtained the right medial prefrontal cortical activity for nonverbal ToM tests in which the participants had to select an appropriate ending to a pictorial scene based on the character's intentional states.

One fascinating hypothesis is that ToM may have evolved from the ability to recognize animate agents (Frith & Frith, 1999). An object, even if it does not resemble a living creature, can be perceived as alive if it moves in ways characteristic of biological organisms. This perceptual capacity may have been the starting point for inferring intentions from actions. Brain areas that could subsume perception of biological motion and ToM include the right superior temporal sulcus (STS) and the right temporal–parietal areas

Figure 11.4 The Sally–Anne task used to test ToM (after Gallagher, 2000).



Source: Adapted version of Heinz Wimmer and Josef Perner's Sally–Anne Task. Illustration by Jack L. Frazier. Used with permission of SAGE Publications, Inc.

(Blakemore & Decety, 2001; Ruby & Decety, 2001). Patients with right hemisphere damage have difficulty in tasks requiring the appreciation of other people's state of mind (Siegal, Carrington, & Radel, 1996; Winner, Brownell, Happé, Blum, & Pincus, 1998).

The amygdala complex is also involved in ToM. Monkeys with bilateral amygdala lesions manifest social and emotional disturbances similar to those found in autism (Bachevalier, 1994; Bachevalier, Málková, & Mishkin, 2001). These monkeys demonstrate social avoidance, less initiation of social contact, and difficulties related to eye contact, facial expression, and body posture. The amygdala is involved in the fear response, so we would expect damage to these areas to produce deficits in the emotional aspects of social interaction, which is what these findings show.

Siegal and Varley (2002) conclude that the neural machinery underlying ToM reasoning involves both evolved specialized capacity with dedicated circuits and a reliance on existing general cognitive structures. They also argue that experience is necessary for its development. ToM seems related to other abilities, such as language, that rely on learning to come to their full fruition. In particular, early conversational experience can allow the child to acquire knowledge about other minds. It may do this by demonstrating that other people act in accordance with their statements and by allowing the potential for imagination of hypotheticals (see Table 11.3).

Autism

Autism is a neurodevelopmental disorder characterized by impaired social interaction and accompanied by restricted and repetitive behavior (Bailey, Phillips, & Rutter, 1996). These individuals are often aloof during childhood and remain egocentric as adults. They show impairments in verbal and nonverbal communication. Some do not speak

Table 11.3 Brain areas underlying theory of mind reasoning (after Siegal & Varley, 2002).

Brain Region	General Cognitive Function	Specific Theory of Mind Function
Left hemisphere	Linguistic syntax and propositional reasoning	Entertainment of false belief
Frontal lobes	Executive function and problem solving	Suppression of the object's real location in the Sally–Anne task
Right temporal sulcus and right temporal–parietal areas	Perception of biological motion	Inferring intention
Amygdala	Fear response	Emotional aspects of social interaction

at all, while others may show delayed speech development. Even those who become verbally fluent later in life show language comprehension deficits.

Autistic symptoms persist throughout life but vary considerably in their severity. A mild version not accompanied by cognitive and linguistic developmental delays is known as Asperger syndrome. Patients with Asperger syndrome may show few symptoms until early adulthood. The current belief is that the autistic symptoms show a broad spectrum, and in this case, Asperger syndrome need not be labeled as a separate disorder. Therefore, autism and Asperger syndrome are treated as belonging to the category of autism spectrum disorders.

The overall prevalence of autism in the general population is estimated to be between 0.3% and 0.7%. Males are three times more likely than females to have it. The number of people diagnosed has increased tremendously over the past two decades, although it remains to be seen whether there is a real increase in the number of cases or if there is simply increased awareness and testing (Newschaffer et al., 2007).

Currently, there is no "cure" or preventative measure that can be taken. However, individuals can receive compensatory cognitive and behavioral training and can learn to overcome some of the deficits. Since severity can vary and there are large individual differences, training is often tailored to the needs of the particular individual. The effectiveness of different treatments is difficult to assess due to methodological differences (Ospina et al., 2008).

Autism has clear genetic origins, and susceptibility genes have been identified (Maestrini, Paul, Monaco, & Bailey, 2000). It is not known whether it is caused by genetic mutation or by combinations of genetic variants (Abrahams & Geschwind, 2008). Several environmental causes have been proposed, including agents linked to birth defects, heavy metals, and pesticides (Arndt, Stodgell, & Rodier, 2005; Rutter, 2005).

Autism and ToM

Frith (2001) reviews the literature and shows that one of the key deficits of autism is a failure to mentalize (i.e., "mind blindness"). This is a failure to understand that others have mental states such as beliefs and desires. We have already examined this ability and called it a ToM. A problem with forming ToM could explain a great many autistic symptoms, especially the social interaction and communication difficulties. However, it cannot be used to account for other autistic symptoms such as repetitive behavior and motor problems, which are better explained by other theories.

Having said this, there is ample evidence that autistic children fail to attain or are delayed in attaining certain developmental ToM milestones. They fail to follow another person's gaze, do not point toward or show objects of interest, and do not understand make-believe play (Baron-Cohen, Cox, Baird, Swettenham, & Nightingale, 1996). These first two behaviors are indicators of shared attention. Also, preschool-age autistic children do not show a preference for speech over other auditory stimuli (Klin, 1991), nor do they show a preference for faces over other visual stimuli, such as hats (Hobson, 1993)—whereas typically developing children show a clear preference for speech and for faces.

A large amount of work in neuroscience also points to a failure of ToM ability in autistics. In one study, a group of adults with Asperger syndrome and a second group of

control adults were compared using positron emission tomography imaging while they performed a mentalizing task. Both groups scored equivalently, but the patients with Asperger syndrome showed less activation in the mPFC. They instead showed more peak activation in the ventral frontal cortex (Happé et al., 1996). In another imaging study employing fMRI, autistic and control adults were asked to judge people's emotional states based on facial-area photographs. The autistic participants showed less frontal lobe activation and no amygdala activation in comparison with the controls.

Anatomical studies point to the same conclusion. Compared to control patients, high-functioning autistics have less gray matter (tissue corresponding to groups of neuron cell bodies) in the anterior parts of a system involving the amygdala, the paracingulate sulcus, and the inferior frontal gyrus (Abell et al., 1999). Other studies have shown an increase in gray matter in the peri-amygdaloid cortex, the middle temporal, and the inferior temporal gyrus. In a postmortem examination of autistic brains, Bauman and Kemper (2007) discovered smaller neurons and a greater density of neurons in parts of the limbic system. Outside this region, they found fewer Purkinje cells in the posterior and inferior cerebellum.

OTHER SOCIAL COGNITIVE DISORDERS

A fascinating but underinvestigated social cognitive disorder goes by the name of **Williams syndrome**. This is a genetic disease caused by a deletion on chromosome 7. These individuals manifest hypersociability. They show an exaggerated interest in other people and increased expressiveness and social communicability. For instance, when presented with a picture showing a social scene, they describe it in greater detail and complexity than do a control group of children with Down syndrome (Bellugi, Lichtenberger, Jones, Lai, & St. George, 2000). Unfortunately, Williams syndrome patients are deficient in other areas of cognition. They have limited spatial skills and motor control and experience learning disabilities. They also have heart, kidney, hearing, and musculoskeletal problems.

A different, more common, social cognitive disorder is **social phobia**, characterized by fear of public places, social interaction, and being evaluated negatively by others (Baumeister & Leary, 1995). In several studies, social phobics were found to have greater amygdala activation when viewing neutral or angry faces and when getting ready to give a public speech (Birbaumer et al., 1998; Tillfors et al., 2001). Clearly, more research needs to be done on these disorders and with people who have greater than average social skills (Adolphs, 2003).

ATTITUDES

An **attitude** is a learned predisposition to respond to a particular object in a particular way (Petty & Wegener, 1998). *Object* is meant in the broad sense here to include physical objects, such as an automobile; people, such as your cousin; and ideas, such as abortion. Attitudes consist of three components. They can be (1) cognitive, consisting of beliefs

and thoughts; (2) affective, referring to feelings and emotions; or (3) behavioral, meaning they influence the way in which we act. For instance, you may have a negative attitude toward your roommate. This might cause you to think that they are inconsiderate, to feel anger toward them, and then tell them to turn down their music.

It is natural to assume that our attitudes affect our actions. After all, if your attitude toward ice cream is positive, you are likely to buy and eat it. Research shows, however, that it is possible to think one way about something but to act in another way toward it. Davis, Grover, Becker, and McGregor (1992) surveyed college students and found that more than 90% of them thought that cheating was wrong. In contrast to this, 40% to 60% of these very same students reported that they had cheated on at least one test (Davis & Ludvigson, 1995).

So when do our attitudes guide our actions? A number of studies have found that at least three factors exert an influence (Kraus, 1995). To begin, they found that attitudes do guide actions when outside influences are minimal. For example, you might believe that the federal government should not have bailed out failing banks but will resist saying so in the presence of your friend who works for a bank. Attitudes also guide action when they are more specifically relevant to a behavior. Your uncle may claim to value good health (a general attitude) while at the same time eating junk food and drinking lots of beer. But his positive attitude toward working out at the gym (a specific attitude toward a behavior) is more likely to influence how often he goes to the health club to work out.

The third factor concerns how aware we are of our attitudes. Attitudes are more likely to influence our behavior the more conscious we are of them. Fazio (1990) found that the attitudes that come to mind easily are those that are most likely to guide our actions. Sarah spends a lot of time thinking about drug legalization. She used to smoke marijuana when she was younger and now works for an advocacy group trying to convince the state legislature to legalize the drug. Her attitude toward marijuana is thus very strong. It comes to her mind easily and should have a powerful influence on her behavior. If there is a march to legalize marijuana in her town, she will in all likelihood march in it.

We can now ask the reverse question. Do our behaviors influence our attitudes? Suppose you didn't want to do volunteer work in your town but had to. Would you end up changing your attitude toward it? If so, would you like it more or less? Research indicates that when people are asked to repeatedly assert an attitude, they are more likely to act in a fashion consistent with it (Powell & Fazio, 1984). So even just stating an attitude may get us to behave as if we possess it. This process can be summed up as "saying is believing."

According to the **foot-in-the-door phenomenon**, people who agree to a small request now may comply with a larger one later. Numerous studies show that people coerced into doing something against their will later change their mind and tend to agree with the attitude behind the new behavior more (Cialdini, 1993). In other words, "doing" now becomes "believing" later. American prisoners of war in Korea were asked to perform small tasks against their beliefs, such as speaking or writing out trivial statements. They then escalated to behaviors even more against their attitudes, such as publicly condemning the United States. This process was colloquially referred to as "brainwashing."

In fact, one can even be fooled by a magic trick into changing one's attitude on a moral dilemma. Lars Hall and colleagues (Hall, Johansson, & Strandberg, 2012; Pärnamets, Hall, & Johansson, 2015) developed a "choice blindness" paradigm where

people complete a moral/political survey on a clipboard, but when they hand it back to the experimenter, he uses a sleight-of-hand trick to replace some of their responses with different answers. Next, the participant is asked to look at their (altered) answers and explain them. The majority of participants did not detect that some of their answers had been altered and, in some cases, even coherently defended the new moral/political positions they had been tricked into endorsing.

Cognitive Dissonance

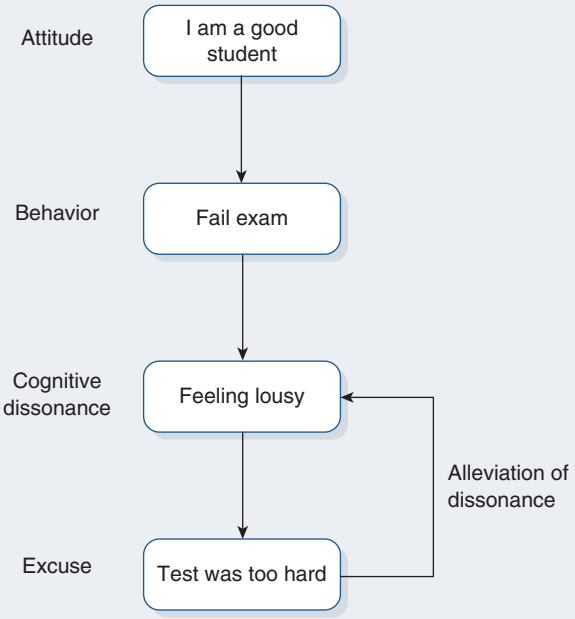
In this section, we will explore what happens when there is an inconsistency between our attitudes and our behavior. Most of us like to think of ourselves as consistent. We believe that the ways in which we think, feel, and act are the same. A contradiction between our beliefs or between our beliefs and actions might then make us feel bad. Here is a case in point: Joe is proud of the fact that he is responsible and on time. He demonstrates this by picking up his friend Jessica to drive her to work at exactly 8:00 a.m. every morning. This Monday, for the first time in years, he arrived 15 minutes late. How do you think Joe might feel? If Joe's negative state resulted from a mismatch between his self-perception and his actions, we say that he is experiencing cognitive dissonance. **Cognitive dissonance** is the uncomfortable psychological state that occurs when an attitude and a behavior (or two attitudes) are inconsistent with each other. Losch and Cacioppo (1990) found that this state is accompanied by a heightened level of arousal.

Since cognitive dissonance doesn't feel good, how might we go about reducing it? In Joe's case, he may make up an excuse, saying he stayed up too late the night before or that his alarm clock failed to go off. In this manner, he can treat the event as an exception and preserve the consistency between his attitude about himself and his behavior. This is an example of a direct strategy because it involves actually changing our attitudes or behavior. Another example of this type is trivializing the incident—that is, considering it as unimportant (Simon, Greenberg, & Brehm, 1995). Alternatively, we could employ an indirect strategy, such as trying to feel good about ourselves in other areas of life (see Figure 11.5).

Does the reduction of cognitive dissonance require conscious reasoning? For this to happen, a person may have to rationalize his or her choices after the fact. That is, a person would have to justify to himself or herself that the decision was a good one. If this is the case, then amnesiacs who forget this thought process may not experience dissonance reduction. To test this, Lieberman, Ochsner, Gilbert, and Schacter (2001) created cognitive dissonance by forcing amnesiacs to choose one set of stimuli from two sets they liked equally. They eventually came to like the selected one more and the rejected one less, the same result obtained with nonamnesiacs. So, apparently, explicit conscious recollection of the justification process after the choice is not required.

In a second experiment, these authors imposed a cognitive load on participants during the decision-making process itself rather than after it. A cognitive load means having to perform some secondary task that takes up processing resources. This prevents an individual from being able to think too much while performing the primary task, which in this case was the dissonance procedure. Participants in the cognitive load condition and those in another condition without a load showed dissonance reduction. These results suggest that attitude change can be either an automatic process that doesn't require full attentional awareness or a controlled cognitive one.

Figure 11.5 Steps in the process of cognitive dissonance.



Attitudes and Cognitive Processes

A large body of research exists on the role of dissonance in different cognitive tasks. In this literature, investigators study consistency judgments. They look to see whether participants will respond in ways that favor similarity between their attitudes or between their attitudes and their behavior. Fiske and Taylor (2008) divide the literature into five overall processes: (1) perception, (2) attention, (3) interpretation, (4) learning, and (5) memory. We follow this organization and report these findings in the sections that follow.

Perception

According to the principle of **selective exposure**, people will seek information not already present that supports consistency. There is some evidence that we seek out information suited to our needs and dispositions (Sears, 1968). For instance, the types of people we choose as friends, the sorts of movies we like to watch, and the music we listen to can all reinforce our attitudes. Other people we know who also share those interests can then reinforce these attitudes. However, there is some controversy in the literature over the validity of this phenomenon. Information that is useful or new can sometimes override dissonance in determining exposure (Brock, Albert, & Becker, 1970).

Attention

The selective attention view proposes that we notice or pay attention to consistent information once it is present. The evidence supporting this view is stronger than that for selective exposure. Studies show that people will spend more time looking at consistent rather than inconsistent information (Brock & Balloun, 1967). In one experiment, participants were divided into two categories (Olson & Zanna, 1979). The repressors were people who usually avoid threatening stimuli, while the sensitizers were those who seek out dangerous stimuli. Each of these groups was shown paintings and then reported their attitudes about them. They next were allowed to keep two paintings—one they liked and another they did not—and the amount of time they spent looking at each was recorded. Only the repressors spent more time gazing at their preferred picture and avoided looking at the inconsistent one. According to cognitive dissonance theory, both groups should have behaved this way.

Interpretation

If cognitive dissonance causes us to focus on consistent information, then we can also make a prediction regarding the process of interpretation. We can expect that people will translate or interpret ambiguous information to be consistent with their attitudes. A variety of studies suggest that this is, indeed, the case. Holbrook, Berent, Krosnick, Visser, and Boninger (2005) found that participants' attitudes toward presidential candidates influenced how they subsequently judged their debate performances. Attitudes regarding the conflict between the Israelis and the Palestinians affected how fair news coverage of the topic was perceived to be (Vallone, Ross, & Lepper, 1985). A more recent study showed that Germans' attitudes about the Euro currency held sway over their judgment of inflation (Traut-Mattausch, Schulz-Hardt, Greitemeyer, & Frey, 2004).

Learning

Dissonance theory predicts that we ought to more easily learn material that is consistent with our attitudes. If two people were arguing over abortion, we would expect the prolifer to learn more of the prolife arguments and to forget more of the prochoice arguments. Conversely, the prochoice person ought to remember more of the prochoice arguments and to forget more of the prolife arguments. There is some evidence in support of this view. Malpass (1969) found that people recalled more consistent information under conditions of incidental learning when they didn't know that they would be tested on the material. Under intentional learning conditions, both types of information were acquired equally. Individuals who were highly motivated to learn also failed to show any consistency effect (Zanna & Olson, 1982).

Memory

Consistency effects are also found for memory. In one study, participants were asked to judge how suitable various people were for different jobs (Lingle, Geva, Ostrom, Leippe, & Baumgardner, 1979). If a person was judged as good for a particular job, then more of those particular job-relevant characteristics were recalled about them later. For instance, if a person judged another person as being a good comedian, then that person

would be more likely to recall the other person's infectious laughter than his or her eye color.

Interestingly, it is possible to hold two attitudes about something simultaneously in memory. Under dual-attitude conditions, one holds an older belief that was formed initially and a newer different belief (Wilson, Lindsey, & Schooler, 2000). The more recent attitude exerts more influence in conscious voluntary cognitive processing, while the older attitude may come into play when one is asked to respond quickly or under cognitive load (having to hold more information in memory or to solve more difficult problems). The two attitudes can also interact to influence consistency judgments (Petty, Tormala, Brinol, & Jarvis, 2006).

Attitudes and Neuroscience

An *evaluative* categorization is one that implies better or worse. It requires that a person think of one thing as positive and another as negative. If one held stereotypically negative attitudes of Blacks compared with whites, then categorizing whites and Blacks would be evaluative. In contrast, a nonevaluative categorization requires a division based on conceptual or semantic criteria that are not value based. An example of a nonevaluative judgment would involve deciding whether items are either a vegetable or a nonvegetable.

Cacioppo, Crites, and Gardner (1996) examined these two types of judgments in participants while recording neural responses via event-related potentials. They discovered posterior right hemisphere activation for evaluative categorizations and activity in both hemispheres for nonevaluative categorizations. The results suggest that different, although not completely independent, neural systems subsume each kind of judgment. Evaluative categorization may be a fast automatic process that has evolved to "kick in" quickly to assess whether a stimulus poses a threat. Nonevaluative judgments appear to be more controlled and slower and to draw on left hemisphere centers involved in semantic and associative meaning.

As we have mentioned earlier, the amygdala plays an important role in the fearful response to threatening stimuli. This brain structure seems to be involved in attitude expression as well. Adolphs and Tranel (1999) found that damage to the amygdala produced more positive preference ratings for pictorial stimuli. One interpretation of these results is that stimuli normally elicit both positive and negative responses but that the amygdala seems to contribute primarily to negative ones. When this structure is damaged, these responses are diminished, enhancing the positive evaluation (see Table 11.4).

Table 11.4 Neural areas and cognitive factors underlying attitude processing (after Cacioppo, Crites, & Gardner, 1996).

Categorization Type	Brain Area	Cognitive Process	Cognitive Function
Evaluative	Posterior right hemisphere (and amygdala)	Fast and automatic	Threat assessment
Nonevaluative	Right and left hemispheres	Controlled and slow	Semantic and associative processing

IMPRESSIONS

An **impression** is the opinion or belief that you form about another person. Whereas an attitude is general and can be formed about an object, a person, or an idea, an impression is formed toward a person only. A number of studies show that early impressions are very powerful (Schlenker, 1980). In one experiment, participants were asked to watch short 10-second video clips of teachers interacting and speaking with students. Based on this brief evidence alone, they were able to accurately assess what the teachers thought about the students (Babad, Bernieri, & Rosenthal, 1989). In another experiment, participants listened to 20-second audio recordings of medical doctors speaking to patients during an office visit. These listeners were able to predict which of the doctors were most likely to have been sued in the past, simply from the sound of their voices in the recordings (Ambady, Koo, Rosenthal, & Winograd, 2002). These examples show that we very quickly form impressions and that, in some instances, these impressions are accurate.

Research has shown that several different perceivable aspects of people influence our judgment of them. First is their physical appearance. Do we judge people who are attractive more positively? The answer is yes. Our first impressions of attractive people are more favorable than those of people we find less appealing (Dion & Stein, 1978; Feingold, 1991). Attractiveness is most powerful when people have little or no information about one another. It also seems to influence some traits more than others. Attractive individuals are judged to be more sociable but not to have more integrity, modesty, or concern for others. One explanation for this is the **halo effect**. This states that we tend to see positive traits grouped together. If a person is good-looking, we tend to think that he or she is also successful. If a person is intelligent, we tend to think that he or she is also wealthy, and so on.

How people talk also makes a difference. People who speak in a clear and direct manner are judged better than those who use filler words and hesitations such as *like*, *sort of,* and *you know* (Erickson, Lind, Johnson, & O'Barr, 1978). Not just how you talk but what you talk about is also important for impression formation. People who reveal some confidential aspect of themselves, what is called **self-disclosure**, are judged well. However, if a person becomes too self-disclosing too quickly in a relationship, it can be a turnoff, producing a negative impression (Collins & Miller, 1994).

Prior knowledge of a person affects how we think about him or her. In one study, half of the students in a class were asked to read a description of a visiting lecturer that portrayed him as "warm." The other half read a description that labeled him as "cold." Predictably, the students who had read the warm description had a better impression than did those who read the cold description (Kelley, 1950). This and other work have shown that the labels we apply to people have a powerful influence on how we think and act toward them.

The Dual-Process Model of Impression Formation

How do we cognitively go about forming impressions of others? According to the dual-process model, we do it in two stages (Brewer & Harasty Feinstein, 1999; Fiske,

Lin, & Neuberg, 1999). The first stage is fast and automatic and based on categories of people. For instance, imagine you are in a grocery store parking lot and you notice a girl scout walking toward you with her collection of cookie boxes. You might initially place her in the category of "salespersons you don't feel like talking to right now." That initial process forms an impression of her based on that category. Your response might be to try to avoid this girl scout. However, as she walks closer, you might recognize her as the responsible young lady in your neighborhood who helps out some elderly neighbors with their gardening. A slower, more deliberative, stage then takes over in which an impression is formed based on this individual person, not on some broad category to which she might belong. Realizing that she, and her girl scout troop, are likely to do good things with the proceeds from the cookie sales, you might then decide not to avoid her but instead to purchase some cookies from her.

ATTRIBUTION

Johnny the third grader is in the cafeteria eating lunch with his friends. Suddenly, he picks up his food and throws it across the table, hitting Joey in the face. What is the cause of this action? Our attempt to explain this kind of behavior is called an attribution. An **attribution** is an explanation we give for the causes of events or behaviors (Forsterling, 2001). There are two basic types of attributions. An internal attribution explains a person's actions in terms of that person's own emotional state, preferences, or beliefs. In the case of Johnny, we could say that he threw the food because he is a bully. An external attribution explains a person's behavior in terms of the environment or situation. In this example, we might reason that Joey is the bully and was making fun of Johnny, who then retaliated. It may be easy to confuse attributions with attitudes reviewed in the previous section. Here is the difference. Attitudes are general, whereas attributions are very specific, referring only to our perception of the cause of an action or behavior.

What determines whether we make an internal or external attribution? Harold Kelley (1972) proposes several primary factors. One is consistency. If a person regularly reacts in a certain way, we are more likely to attribute that type of action to him or her. Imagine that you hear your friend Mary laughing. If Mary laughs all the time, you will attribute her laughter to her sense of humor. If she rarely laughs, then you may think that her current laughter is due to somebody telling a particularly good joke. Another influence is consensus. According to this view, if lots of other people agree on a behavior, then it is more likely to be attributed to the situation. If you hear your roommate next door cheering and realize that the World Series is on and that most people in your dorm like the local team, then you will attribute the cheer to the team having scored a run. If there is no such consensus, you may simply attribute the cheer to your roommate being a diehard sports fan.

Distinctiveness also plays a role. If a person acts the same way in different situations, then distinctiveness is low. If he or she acts differently in different situations, then distinctiveness is high. Low distinctiveness is like high consistency and produces internal attributions. High distinctiveness instead biases a situational attribution. Under these conditions, we tend to think that the environment or context is responsible. If Courtney

practically always picks chocolate ice cream but tonight chooses pistachio, then we may attribute that choice to the situation (e.g., it must be really good ice cream or, perhaps, the only flavor available).

Attribution Biases

Are our attributions accurate? We hope that they are; otherwise, it would be difficult to understand the world around us and to interact successfully with others. Researchers have found that in some cases, we are not so good at making correct attributions. One of the biggest stumbling blocks is the **fundamental attribution error**. This is the tendency to prefer internal or dispositional traits as the best explanation. In other words, we are more likely to blame a person than a situation for some behavior (Gawronski, 2003; O'Sullivan, 2003). If you are driving home from work and somebody behind you is tailgating, you are likely to believe that it is because that driver is aggressive or in a hurry. Under these conditions, you would tend to ignore contextual information such as the fact that this person is also being tailgated and has sped up to allow more room behind his or her car. Don't confuse cognitive dissonance with the fundamental attribution error. Cognitive dissonance applies primarily to yourself and the feelings you experience during a contradiction between an attitude and a behavior. The fundamental attribution error is about how you explain somebody else's behavior (see Figure 11.6).

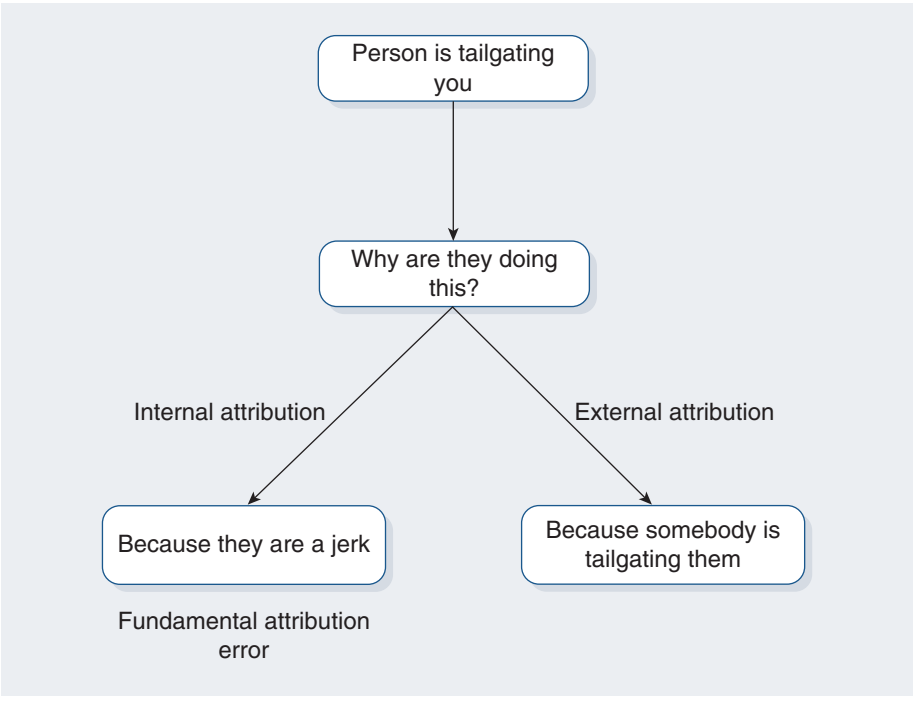
Our attributions can also be distorted by the **self-serving bias** (Brown & Rogers, 1991). This is the tendency to explain our own failures to external causes but to explain other people's failures to internal ones. The reason you did poorly on the test was because you didn't sleep well the night before, but the reason Alexander didn't do well on the test was because he is stupid. The reverse holds for successes. We explain our own successes in terms of internal causes but others' in terms of externals. If you did well on the exam, it was because you are smart, but if Alexander did well on the exam, it was because the test was easy.

According to the **belief-in-a-just-world phenomenon**, we think that people get what they deserve (Lerner, 1980). If we are rich and successful, we think that it is because we are clever or hardworking. Conversely, if we see a beggar on the street, we feel he or she is poor because he or she has a psychological disorder or a substance abuse problem. In other words, we put the cart before the horse, using the end state as the explanation for the cause. It could just as easily be the case that you are successful because you have wealthy parents and the beggar is poor because somebody stole all of his or her money.

Attribution and Cognitive Processes

How do we actually go about making attributions? A number of theories have been proposed that incorporate both fast, automatic attentional mechanisms and slower, more controlled ones (Fiske & Taylor, 2008). There is sufficient evidence to argue that we rapidly and subconsciously form dispositional attributions. In one study, participants were able to make trait judgments based on stimulus features such as gender, race, age,

Figure 11.6 An example of how an attribution is made.



and facial appearance in as little as 100 milliseconds (Willis & Todorov, 2006). These judgments were correlated with those made under longer-duration viewing conditions, suggesting that our first fast impressions tend to agree with our longer, more contemplative ones.

However, controlled attentional processes sometimes do adjust or change our initial impressions. Imagine that you are in a bar and see your friend Mark in a fight. Under these conditions, you may very quickly perform a dispositional assessment of the scene, remembering that Mark is aggressive and has gotten into fights before. This might first lead you to believe that he started the fight. A short time later, you may adjust your initial assessment by noting the situation. Realizing it is a bar, you may then infer that some of the other drunk and unruly patrons could have started the altercation. Gilbert (1998) summarizes older cognitive attribution theories and incorporates some of their features into a multistage model of how it might occur.

Attribution and Neuroscience

Recent work has been able to identify those brain regions involved in making judgments about other people's behavior. These regions consist of the anterior paracingulate cortex,

the posterior STS at the TPJ, and the temporal pole (Gallagher & Frith, 2003). The dorsal mPFC is implicated in the general comprehension of other people's minds. The mPFC plays a role in thinking about our own states as well as those of others. The TPJ in the right hemisphere becomes active for the attribution of other mental states but not for other kinds of social information about them (Rilling, Sanfey, Aronson, Nystrom, & Cohen, 2004).

One particularly interesting finding is that there may be different brain areas that operate when we are dealing with abstract situations involving people (e.g., a fictional scenario or story), in contrast to real-world people scenarios (e.g., playing a game with real people who are in close physical proximity). Rilling et al. (2004) found activity in the anterior paracingulate cortex and the posterior STS when participants were thinking about actual people. There was less activation in these areas for hypothetical situations (Table 11.5).

Harris, Todorov, and Fiske (2005) had a group of participants reason about a scenario with different combinations of consistency, consensus, and distinctiveness. High consistency, low consensus, and low distinctiveness produced greater activation of the STS. High consistency and low distinctiveness, regardless of consensus, produced greater activation of the mPFC. The results suggest that people will tend to ignore consensus information in favor of consistency and distinctiveness.

Lieberman, Gaunt, Gilbert, and Trope (2002) postulate the existence of two neural networks that underlie automatic and controlled processing. The reflexive or X system produces fast automatic responses and is mediated by the amygdala, the dorsal anterior cingulate, the basal ganglia, the ventromedial PFC, and the lateral temporal cortex. This system is used under normal nonconflicting conditions. The reflective or C system takes longer and involves the lateral PFC, the mPFC, the rostral anterior cingulate, the posterior parietal cortex, and the medial temporal lobe. The C system is used when a conflict occurs or our goals are blocked and we need to think more carefully (Table 11.6).

Table 11.5 Neural centers and associated cognitive functions underlying attribution processing.

Brain Area(s)	Cognitive Process
Dorsal medial prefrontal cortex	Thinking about mental states, our own and others'
Temporoparietal junction in the right hemisphere	Attribution of others' mental states
Anterior paracingulate cortex and posterior superior temporal sulcus	Thinking about real-world people scenarios

Table 11.6 Two hypothesized brain systems underlying attribution processing (after Lieberman, Gaunt, Gilbert, & Trope, 2002).

System	Response	Brain Regions	Conditions Used Under
Reflexive (X system)	Fast and automatic	Amygdala Dorsal anterior cingulate Basal ganglia Ventromedial prefrontal cortex Lateral temporal cortex	Normal and nonconflicting
Reflective (C system)	Slow and controlled	Lateral prefrontal cortex Medial prefrontal cortex Rostral anterior cingulate Posterior parietal cortex Medial temporal lobe	Conflict or goal blocked

INTERDISCIPLINARY CROSSROADS: GAME THEORY AND THE PRISONER'S DILEMMA

A number of decision-making situations pit the narrow self-interests of individuals against the interests of the group as a whole. Take public radio as an example. Listeners fund public radio. If you wanted to "cheat" by not paying and rely on other people to support the station for you, you would benefit at their expense. If everybody did this, though, the radio station would go under and nobody could listen. On the other hand, if everyone cooperates, the group as a whole will be better off since the station will continue to stay on the air. The field in the social sciences that studies these kinds of situations, where an individual's success in making choices depends on the choices of others, is known as **game theory**.

These types of situations are called **social dilemmas**. In the simple version, each person is better off doing one sort of thing, but everyone is better off if they all do something else (Baron, 2000). If individuals act selfishly, thinking only of themselves and not doing what the majority does, they are said to be defectors. If people act altruistically, taking other people's concerns into consideration, they become cooperators. The **prisoner's dilemma** is one of the simple kinds of conflict situations such as this that can arise. It portrays the logical consequences of being selfish or altruistic and is a demonstration of the sorts of social factors that can affect the attributions we make toward others.

(Continued)

(Continued)

Imagine that you and your friend have both been charged for a crime and are now waiting in prison to be tried. There are four possible situations that can occur (see Table 11.7). If neither of you confesses (i.e., if the two of you "cooperate" with each other), you each serve minor sentences of 1 year. If you cooperate by not confessing and your partner "defects" by turning in evidence against you, he goes free while you end up serving a 10-year prison sentence. Conversely, if your partner cooperates by not snitching on you and you defect by turning in evidence against him, you get to go free while he serves the 10 years. If both of you defect, you each serve 4 years in jail.

In this scenario, you are each being interrogated separately and so have no idea what the other will do. What would you do? Most people reason that no matter what their friend does, they would serve less time by defecting (0 years if he cooperates or 4 years if he defects) than by cooperating (1 year if he cooperates or 10 if he defects). So the typical outcome is that both defect and each serves 4 years. This is unfortunate for them, since each would have served only 1 year if they both had cooperated. When the prisoner's dilemma is played out in the laboratory, the majority of people end up selfishly defecting.

The type of conflict depicted in the prisoner's dilemma comes up all the time in everyday life, in situations ranging from business contracts to environmental agreements. In fact, these situations are the foundation of all economic systems. If you as an individual decided to be selfish, you might take as much as you could from everyone else while working the least—that is, contributing the least to society in terms of the goods and services provided by your labors. Obviously, not everyone does this, or we would have nothing to enjoy. The problem is solved in most economies through the use of money, for which one needs to work in order to buy things.

Despite this, the default behavior that we see in most real-world scenarios is cooperation, not selfishness. How can this be, given the benefits that reward individualistic action? Well, in the real world, it is unlikely that we will interact with a person on just one occasion. Since we live in a society, the chances are that you will meet your friend again someday. If you screw him over the first time, you realize that he may get back at you the second time. It is this scenario that is studied in the iterated prisoner's dilemma. Here, people can "play" this game with the same person again and again.

Under iterated circumstances, a new strategy develops, called "tit for tat" (Myerson, 1997). People in these situations do whatever their opponent did to them on the previous round. If your friend defected on you previously, you

Table 11.7 The four outcomes in the prisoner's dilemma scenario.

	Your Partner Confesses	Your Partner Cooperates
You confess	Both serve 4-year sentences	You go free; partner serves 10 years
You cooperate	Partner goes free; you serve a 10-year sentence	Both serve 1-year sentences

would then defect on him the next time. If your friend decided to get "nice" at some point and cooperate, then you would reciprocate again by "forgiving" his previous transgressions and cooperating. Results from group-based tournaments of the iterated prisoner's dilemma game suggest that a strategy of "tit for tat" with some occasional forgiveness may be the best approach (Axelrod, 1980). In computer models of many simulated agents interacting via prisoner's dilemma–style exchanges, various circumstances can cause defection to spread like a contagion or for cooperation to spread (Vanderschraaf, 2006). The cultural evolution of cooperation in these agent-based model simulations is influenced by a variety of parameters, such as subgroup affiliation, migration patterns, and resource limitations (Smaldino, Schank, & McElreath, 2013). In realistic scenarios where we have to interact repeatedly with many individuals who can get back at us, it usually makes more sense to be cooperative.

STEREOTYPES

A **schema**, as we have mentioned elsewhere, is a framework or body of knowledge on some topic. Schemas typically help us understand and think about things. For example, you may have a schema about a bicycle that includes an understanding of its parts, such as a seat, handlebars, and wheels. Part of your bicycle schema may also involve an understanding of how they function. This aspect of the schema would include pedaling, gears, and steering.

Schemas need not be used for objects alone. We also use them when thinking about people. One kind of schema is a stereotype. **Stereotypes** contain information about people in a specific category that is oversimplified, exaggerated, or even wholly inaccurate. The category can be defined by almost any criteria, such as skin color, sex, profession, or religion. Stereotypes can be positive, negative, or neutral. For instance, a person who embraces stereotypes might believe that white males are ambitious (positive), entitled (negative), or usually of average intelligence (neutral).

Stereotypes and Cognitive Processes

According to one view, we form schemas to simplify the vast amount of social information with which we are constantly bombarded (Kulik, 2005). In a typical day, you may interact with fellow students, coworkers, friends, and family members. In the course of these interactions, you need to keep track of what each one has said, believes, and intends to do. A stereotype simplifies and organizes this information and makes it easier to think about. Of course, in the process of simplifying, stereotypes may also be wrong and cause us to think incorrectly about people.

Hugenberg and Bodenhausen (2004) found that people prefer information that is consistent with their stereotypes and that they process such information more quickly. We tend not to pay attention to or remember information that fails to match our stereotypes, and in some cases, we believe that this information is false (Johnston & Macrae, 1994;

O'Sullivan & Durso, 1984). To illustrate, if someone held the (incorrect) stereotype that women are bad at math, that person would be less likely to notice or recall a woman who is good at math and instead be more likely to attend to and remember a woman who isn't.

Ingroups and Outgroups

One consequence of stereotyping is that we tend to classify ourselves as part of a preferred category of people, called an ingroup, and others as part of a less preferred category, or outgroup. People perceive more diversity and variation among members of the ingroup and less among those of the outgroup (Guinote & Fiske, 2003). If Frank, who is Caucasian, thinks (incorrectly) that all Asian people look alike, then he has succumbed to the **outgroup homogeneity effect**. People also tend to think of ingroup members as having better looks and personalities and as engaging in more socially accepted forms of behavior, what is known as **ingroup favoritism** (Aboud, 2003; Bennett et al., 2004).

A number of factors influence ingroup favoritism. It occurs even when all the members are deprived and have little to offer one another (Rabbie & Horwitz, 1969). This finding suggests that there is more to it than simply being rewarded by like group members. Ingroup favoritism also increases under conflict, which suggests that it helps reduce subjective uncertainty and anxiety about one's identity.

Outgroup homogeneity is stronger for groups that are unfamiliar and with which one has not had personal contact (Brewer & Brown, 1998; Brown, 2000). Thus, we are more likely to perceive the (incorrect) negative stereotype that Albanians are corrupt if we have never met one. People also believe that others are more biased against outgroups than they themselves are (Judd, Park, Yzerbyt, Gordijn, & Muller, 2005). Interestingly, under some conditions, there is also an ingroup homogeneity effect. This occurs particularly for identity-relevant traits (Simon, 1992). Those who identify strongly with their group are more likely to see members of the group as being alike (Castano & Yzerbyt, 1998).

Automatic Stereotyping

In automatic stereotyping, we associate certain values or traits with particular groups often without being aware of it. The best-known test of these associations is the **implicit association test** (Greenwald, McGhee, & Schwartz, 1998) where participants go through a sequence of tasks that test the strength of their association between a given concept and the value of "good" or the value of "bad" by measuring how quickly they respond in their button presses to those stimuli. A faster response suggests a stronger association between the two stimuli. For example, in a related implicit measurement task, Fazio, Jackson, Dunton, and Williams (1995) had participants judge whether a word that appeared in the middle of a computer screen (e.g., *wonderful* or *disgusting*) was a positive word or a negative word, and their reaction time was recorded. However, shortly before that word was presented, the face of a white person or a Black person was briefly flashed on the screen. Fazio et al. found that their white participants pressed the positive-word button faster when that word had been preceded by a white person's face than when it had been preceded by a Black person's face. And they pressed the negative-word button faster

when that word had been preceded by a Black person's face than when it had been preceded by a white person's face. Fazio et al. interpreted these results as evidence that these white participants had stronger implicit associations between white people and positive words than between white people and negative words, as well as stronger implicit associations between Black people and negative words than between Black people and positive words. Similar evidence for implicit prejudice was observed by Wojnowicz, Ferguson, Dale, and Spivey (2009). They presented white participants with phrases like "Black people" and "white people" and had them move their mouse cursor to response options in the upper corners of the screen stating "LIKE" and "DISLIKE." All participants explicitly reported that they "LIKE" both groups of people, but their implicit bias against Black people was evident in a greater curvature of the mouse trajectory slightly toward the "DISLIKE" response option when the prompt was "Black people."

A large number of studies have demonstrated what is called the **category confusion effect**. Simply stated, this is the difficulty we have in differentiating people who fall within an outgroup category defined by gender, race, age, sexual orientation, attractiveness, and other features (Maddox & Chase, 2004). Perhaps you have experienced this at a community gathering of some sort where you may have remembered that an elderly person made a statement, but you can't remember specifically *which* elderly person it was. This memory effect is an example of automatic stereotyping because it occurs without any intention or effort on our part.

Stereotyping and Neuroscience

Hart et al. (2000) used fMRI to see how the perception of ingroup and outgroup members differs. In this study, white and Black Americans viewed unfamiliar white and Black faces and were asked to identify their gender. In the first set of trials, there was amygdala activation in both the ingroup and outgroup faces. In the second set of trials, there was reduced amygdala activation but only for the ingroup members. This was true for both the races that were tested. We would expect the amygdala to be active when first viewing unfamiliar faces of any race, since this brain structure constitutes part of the fear response and strangers may pose a threat. The reduction in this response for ingroup members only, however, demonstrates that we rapidly get used to faces of our own race, become less fearful of them, and apparently don't consider them to be as much of a threat. The same cannot be said for outgroup members.

In a related experiment, Lieberman, Hariri, and Bookheimer (2001) had Black and white Americans view three faces arranged in a triangular configuration. In the first condition, the task was simply to match which of the two faces on the bottom matched the race of the face at the top. The second condition required deciding which of two labels, "African American" or "Caucasian," applied to a target face. For outgroup target faces, amygdala activation was high in the race-matching condition but low in the race-labeling condition. Measurements of prefrontal activation showed exactly the reverse pattern. For this region, activation was low for race matching and high for race labeling. The results demonstrate that processing of a stereotypical word can reduce our apparent emotional sensitivity to it—perhaps because different, nonstereotypical associations were made while learning the word.

PREJUDICE

A literal translation of the word **prejudice** yields "prejudgment," so we can think of it as a decision or opinion we already have made before collecting any relevant evidence. Myers (2001) describes prejudice as having three components: (1) a belief, such as a stereotype; (2) an emotion, such as envy or fear; and (3) a predisposition to an action, such as discrimination. For instance, Greg may be prejudiced against immigrants, believing that they are violent, feeling fearful that they might rob him, and being unwilling to adopt one as a friend. As is the case with stereotypes, we are less likely to attend to or remember information that refutes or disagrees with our prejudices. Conversely, we are more likely to pay attention to information that supports those stereotypes. The consequence of this is that our prejudices tend to reinforce themselves. The difference between stereotypes and prejudices are that the former are beliefs or thoughts and are thus cognitive in nature, whereas the latter entail not only thoughts but emotions and actions as well.

Again, as we saw with stereotypes, the cognitive component of a prejudice can be a schema that affects what we perceive and understand. In one early study, white participants perceived a white man pushing a Black man as playful, but when these same observers saw a Black man shove a white man, they interpreted the event as aggressive and violent (Duncan, 1976). This phenomenon is a good example of "top-down" or conceptually driven perceptual processing, where our assumptions influence how we interpret what we see. To find out more, you can refer to Chapter 4, where we first introduced this notion.

Is prejudice conscious or unconscious? If it is conscious, we might be able to control and perhaps stop it. If it is unconscious, it might influence our thoughts and actions in ways we can't control. The evidence suggests that prejudice can operate on both levels (Carter & Rice, 1997). An explicit bad feeling toward someone may contribute to the formation of a negative conscious prejudice against that person. However, even the absence of a good feeling can lead to an unconscious negative prejudice (Pettigrew & Meertens, 1995). In this latter case, we could develop a bad prejudice toward an outgroup, even though we have never met an individual member of that group. Using the implicit association test, Greenwald et al. (1998) showed that people who claimed not to be racist took more time to identify positive words such as *paradise* and *peace* when they were associated with Black instead of white faces. So although we may believe that we are not prejudiced, we may still harbor unconscious beliefs that can affect our behavior.

The Stereotype Content Model of Prejudice

In this section, we will describe one model by which we form prejudiced beliefs. According to the stereotype content model, we first adopt particular stereotypical views of some group. These views then determine the emotional reaction or prejudice we feel toward members of that group. There are two primary dimensions by which we judge others. The first is competence, which can be viewed as high or low. The second is warmth, which can be either cold or warm. A number of studies have shown that we tend to classify others using these two dimensions (Fiske, Cuddy, Glick, & Xu, 2002).

Table 11.8 The stereotype content model of prejudice.

	Low Competence	High Competence
High warmth	The elderly and the disabled (viewed by some with pity)	Middle class, Christians, and heterosexuals (viewed most favorably, by some with pride)
Low warmth	Homeless, drug addicts, and the poor (viewed least favorably, by some with disgust)	Jews, Asians, the wealthy, and professionals (viewed by some with envy)

Table 11.8 shows this two-dimensional space and lists several instances of stereotyped people who fit into the four extremes. According to some stereotypes, people who are middle class, Christian, or heterosexual are perceived as being high in competence and warm and so fit into the upper, right-hand portion of the space. These people are viewed most favorably. At the opposite end of the space, in the lower-left corner, there are stereotypes for people who are perceived as low in competence and cold. Homeless individuals, drug addicts, and the poor fit into this area. These groups tend to be viewed least favorably. Both of these groups are seen as unambivalent, meaning that people have clear-cut and strong views of them, either positive or negative.

In the upper-left portion of this space are those perceived as low in competence but high in warmth. Examples include the elderly and the disabled. Finally, in the lower-right corner, we have people seen as cold but high in competence. Jews, Asians, the rich, and professionals fill this category. These two groups are ambivalent because people's views toward them are mixed and uncertain.

By the stereotype content model, each of the four categorizations outlined above elicits a different prejudice and behavior (Cuddy, Fiske, & Glick, 2007). We feel pride toward those warm and high in competence because we usually associate these as members of our own group who can aid us in times of trouble. We are motivated to interact with these groups. Cold, low-competence individuals produce a feeling of contempt, disgust, and dehumanization, motivating us to avoid them. Warm, low-competence groups make us feel pity, which may prompt helping behavior but also neglect. Groups judged as cold and highly competent are seen as a potential threat since they may have needed resources. During stable social times, these groups are cooperated with, but during social upheaval, they can be attacked—as has been the case with genocides throughout history.

OVERALL EVALUATION OF THE SOCIAL APPROACH

You may have noticed a theme that runs through some of the research discussed here. This is the distinction between fast, automatic processing that appears to take place without effort and slower, more controlled, effortful processing. We saw this in the dual model of impression formation. Many social cognitive processes seem to operate

according to this two-track procedure. This echoes what we have seen in other types of cognition. Remember that in visual search, there appears to be a fast parallel detection of targets but also slower serial search. In our discussion of executive function, we saw that there was not only fast automatic processing that could govern well-learned tasks, such as driving a car, but also slower processing for novel or more difficult problems.

Cognitive science's narrow focus on the cognition of the individual has held sway for far too long. Its emphasis on the processing of objects and abstract information such as pictures, numbers, and words was the dominant paradigm for much of the past few decades. Fortunately, we have seen a change in this outlook. As this chapter attests, there is now a substantial body of research on how we cognate about one another. This change is welcome. It helps explain the way we think about and interact with one another, not just with objects and ideas.

The social cognitive perspective has adopted the interdisciplinary approach more readily than many other areas in cognitive science. In fact, by its very definition, it employs at least two levels of analysis, (1) the social and (2) the cognitive. With the current trend of social cognitive neuroscience, we can add a third biological level to this. The intersection of these approaches to the study of cognition should serve as a model for the rest of cognitive science to follow and will continue to yield valuable insights for the foreseeable future.

SUMMING UP: A REVIEW OF CHAPTER 11

1. Social cognition is the study of how we think about ourselves and other people. Researchers in this area make four assumptions: (1) the use of mental representations, (2) how they develop and change over time, (3) an interdisciplinary approach, and (4) real-world applicability.
2. The field of social cognitive neuroscience attempts to explain cognitive phenomena from three perspectives: (1) the biological or brain-based approach; (2) a cognitive, information-processing approach; and (3) a social perspective.
3. Recent research has found that young children outperform primates on social cognition tasks, implying that this capacity may have evolved first and served as the basis for more general cognitive abilities later.
4. Social cognition requires joint attention, the ability to pay attention not just to your own thoughts but to the sharing of an attentional focus with others as a normal part of social interaction.
5. Mirror neurons are brain cells that fire both when an animal is performing some action and when it observes another animal performing that same action. These neurons exist in monkeys and probably form the basis for modeling or imitative behavior. The evidence for the presence of these neurons in humans is not as strong, but if present in humans, they may form the basis for empathy.
6. A ToM is the knowledge that other people have minds and also the ability to use this understanding to infer other people's actions and to adjust our own behavior in response.

Autistic patients seem to have deficits in being able to form and use a ToM, what may be called "mind blindness."

7. An attitude is a learned predisposition to respond to a particular object in a particular way. Cognitive dissonance can arise when an attitude and a behavior are inconsistent. This results in a negative aroused state that can be resolved by changing our attitude to correspond to our actions.
8. An impression is an opinion or belief that you form about another person. Early impressions are powerful. They can be influenced by people's attractiveness, the way they talk, nonverbal body language, and prior knowledge.
9. An attribution is an explanation we give to explain somebody else's behavior. Attributions can be internal, attributed to the person's nature, or external, attributed to the environment. We prefer internal attributions to explain others' behavior but external attributions to explain our own behavior.
10. Stereotypes are a set of ideas or beliefs about a category of people, such as Blacks or Hispanics. Stereotypes cause us to categorize ourselves as members of an ingroup that we perceive as more heterogeneous. Members of an outgroup are perceived as more homogeneous. A prejudice consists of a belief, an emotion, and a predisposition to action. The cognitive component of both stereotypes and prejudices can be schemas.
11. The social cognitive approach again shows us that there are many cognitive processes that have a two-track organization, being fast and automatic or slower and more effortful. This approach is a welcome change because it focuses on how we think about people rather than how we think about objects or ideas.

SUGGESTED READINGS

Banaji, M. R. & Greenwald, A. G. (2016). *Blindspot: Hidden biases of good people*. New York, NY: Bantam Publishing.

Doherty, M. J. (2008). *Theory of mind: How children understand others' thoughts and feelings.* New York, NY: Psychology Press.

Fiske, S. T., & Taylor, S. E. (2008). *Social cognition: From brains to culture.* New York, NY: McGraw-Hill.

Osborne, M. J. (2003). *An introduction to game theory.* Oxford, England: Oxford University Press.

Sicile-Kira, C. (2004). *Autism spectrum disorders: The complete guide to understanding autism, Asperger syndrome, pervasive developmental disorder, and other ASDs.* New York, NY: Perigee Trade.

CHAPTER TWELVE

THE ARTIFICIAL INTELLIGENCE APPROACH I

The Computer as a Cognitive Agent

COGNITIVE SCIENCE AND ARTIFICIAL INTELLIGENCE

We live on a constantly changing planet, and artificial intelligence (AI) is everywhere. The anthropological definition of the current age has been proposed as the Anthropocene (age) by the International Commission on Stratigraphy.[1] The definition is a combination of words: *anthro*, from "man" and *cene* from "new." It reflects the fact this is a period during which humans have had a dominant influence on the environment and world, suggesting that it began with the Trinity Test of 1945, where the use of atomic energy was demonstrated in its most dramatic form. It also subsumes the "knowledge age" in which natural human intelligence is transformed into artificial form that mimics cognitive functions associated with the human mind such as "learning," "problem solving," and "decision making." The agents (both abstract as in an algorithm and mechanical as in a robot) constitute artificial intelligence or AI, which affects all aspects of our daily lives: work, business, social, economic, and educational.[2] A complete history could fill its own text (Nilsson, 2010). While making no attempt to do so, we can view its development from a broad perspective.

[1]As of May 2019, the Anthropocene Working Group (AWG) voted in favor of making a formal naming proposal to the International Commission on Stratigraphy.

[2]A brief list of notable AI accomplishments:

- In the Business Section of the *New York Times* (September 4, 2019), an article is entitled "A Breakthrough for AI Technology: Passing an 8th-Grade Science Test."

Learning Objectives

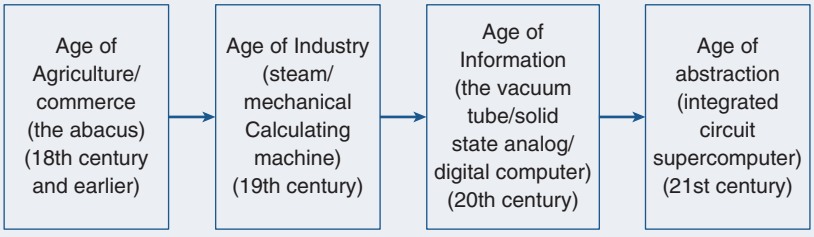
After reading this chapter, you will be able to:

1. Discuss the computational elements of an intelligent agent.
2. Compare symbolic AI to connectionist AI.
3. Describe IF–THEN rules in formal logical AI.
4. Define *fuzzy logic*.
5. Compare deductive, inductive, and abductive reasoning.
6. Describe deep learning algorithms in machine learning.
7. Understand the components of Bayes' theorem.
8. Discuss brain–computer interfaces.

AI has evolved over four distinct epochs, as shown in Figure 12.1. The Age of Abstraction incorporates formal analysis of information as well as human problem solving and decision making. Abstraction deals with ideas as opposed to events. It is characterized by establishing relationships between ideas (e.g., hunger is a sensation produced by the need for food). Information precedes abstraction and provides facts—truth or falsity apprehended by a cognitive agent—or knowledge learned about something of someone (e.g., dendrites are neuronal receptors).

Cognitive science embraces the interdisciplinary study of human characteristics: consciousness, perception, thinking, judgment, and memory, as well as the behaviors that follow. As such, it comprises elements of intelligence (e.g., information representation and processing) and behavior. Such questions as what constitutes reasoning and how is it expressed are primary areas of research and development. Cognitive science and AI

Figure 12.1 Historical development of AI.



- Google's AI can now translate your speech while keeping your voice: *MIT Technology Review*, May 2019.
- Facebook is funding brain experiments to create a device that reads your mind: *MIT Technology Review*, July 2019.
- Instead of practicing, this AI mastered chess by reading about it: *MIT Technology Review*, July 2019.
- "AI and Bionic Eyes Are Helping to Contain Raging Wildfires": Brian Sullivan, August 5, 2019, Bloomberg.com.
- The android Turing Test can tell us whether a robot is effectively a person (https://thenextweb.com/syndication/2019/08/11).

"How High Tech Is Transforming One of the Oldest Jobs: Farming": *New York Times*, September 8, 2019 (https://www.nytimes.com/2019/09/06/business/farming-technology-agriculture.html?smid=nytcore-ios-share).

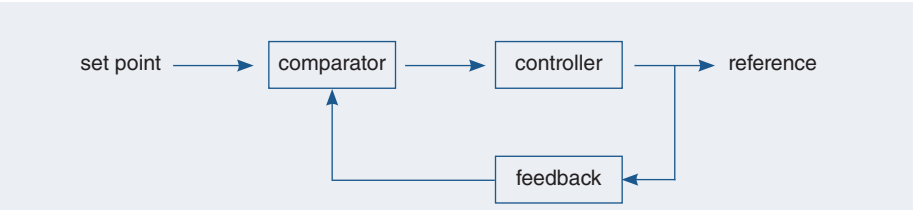
encompass a number of disciplines; a significant subset is shown in Table 12.1. In this and the subsequent chapter, we explore key elements relevant to AI.

Norbert Wiener is recognized as the wellspring of AI. His seminal ideas were first presented in two works: one dealing with cybernetics and a second with the social implications of this technology (Wiener, 1950, 1961). Cybernetics has many definitions, but the simplest one is "the science of communication and control theory" with its comparative study of the nervous system. Wiener wrote at the cusp of an era that changed our perception of machines and humans in physical terms and redefined them as "computational" processes. Changes in such perception are likely to change just as rapidly in the coming eras. Wiener essentially proposed a control system that included a "set point" (a desired outcome), a "reference point" (an achieved outcome), and feedback to compare these variables. If there is a difference, the controller adjusts the reference to reduce the differences. (Physiological control provides a number of cybernetic examples: regulation of blood pressure, cholesterol levels, body movements, and equilibrium.) Wiener's concept is depicted in Figure 12.2. He formalized the analysis of earlier industrial control devices such as the fly wheel and opened a new era of technological advancement as well as its societal implications. In general, it is descriptive of homeostasis.

Table 12.1 Disciplinary contributions to AI.

Classification	Discipline
Formal (physical) sciences	Mathematics and logic, physics, chemistry, earth science, theoretical computer science, information theory, systems theory, decision theory, statistics, theoretical linguistics
Natural sciences	Biology (microbiology, anatomy, neurology, neuroscience, immunology, genetics, physiology, pathology, biophysics, biolinguistics, ophthalmology), zoology
Social sciences	Human behavior and societies: anthropology, archaeology, criminology, economics, public health, sociology, psychology, communication, education, geography, history, law
Applied sciences	Engineering, applied mathematics, applied physics, medicine, computer science, data science
Philosophy	Natural philosophy (processes of transformation in the physical world), moral philosophy (ethics), metaphysical philosophy (logic, existence, causation)

Figure 12.2 Cybernetic system after Norbert Wiener.



The limitations of the Wiener model rest on its inability to account for changing environmental factors as well as restriction to a single goal (set point). The development of digital technology has changed the discussion with the resultant evolution of "AI."

DEFINING AI

In 2016, Stanford University published a report providing an extensive and comprehensive review of AI. The study lasted for a period of 2 years and included a group of world-renowned experts.[3] A primary observation in that report follows:

> Artificial Intelligence (AI) is a science and a set of computational technologies that are inspired by—but typically operate quite differently from—the ways people use their nervous systems and bodies to sense, learn, reason, and take action.

It also includes the following consequential observation:

> Curiously, the lack of a precise, universally accepted definition of AI probably has helped the field to grow, blossom, and advance at an ever-accelerating pace.

(A definition of "human intelligence" is similarly amorphous.) Several contributors to the understanding of AI have provided variations to this meaning. These include the following:

> Nils J. Nilsson: "Artificial intelligence is that activity devoted to making machines intelligent."

[3]https://ai100.stanford.edu/sites/g/files/sbiybj9861/f/ai100report10032016fnl_singles.pdf

Stuart Russell: "a field concerned with pure intelligence, independent of the objective, to a field concerned with systems that are provably beneficial for humans."

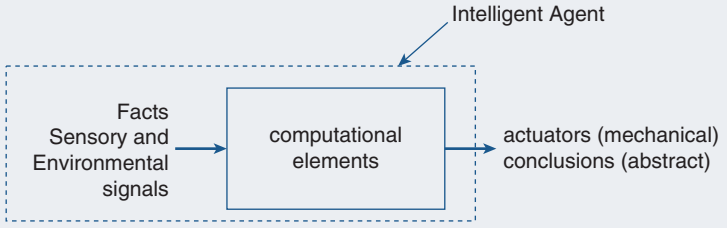
Russell recently modified this view: "Machines are intelligent to the extent that their actions can be expected to achieve their objectives."

We propose to view AI as an "intelligent agent," which is depicted in Figure 12.3.

The inclusion of the term *artificial* may be inadequate as a descriptive of AI. In Chapter 13, the intelligent agent (IA) in Figure 12.3 is shown to be the human brain and a critical element of some applications.

The IA draws its inspiration from the philosophical theory about the relationship between mind and matter. The psychological approach to cognitive science was explored in Chapter 3, and Figure 12.4 mirrors the mind–body concepts as they relate to IAs in the form of "symbolism" and "connectionism." Such agents are concerned with developing machines that solve "real-world" problems (e.g., parking a car without human intervention). A superintelligence is embedded in a machine (computer) with thought and reasoning processes equaling or surpassing the human mind. A singularity is the ultimate development of a superintelligence wherein the machine (superintelligence) enables humanity to transcend its biological limitations; man and machine will be indistinguishable (Kurzweil, 2005).

Figure 12.3 Defining diagram of an intelligent agent.



Facts: predicates having objective reality. A predicate is an affirmation or assertion of a proposition as in the following: "all humans are mortal"

Sensory signals: derivatives of vision, hearing, touch, taste, smell, balance, proprioception, neuronal signals.

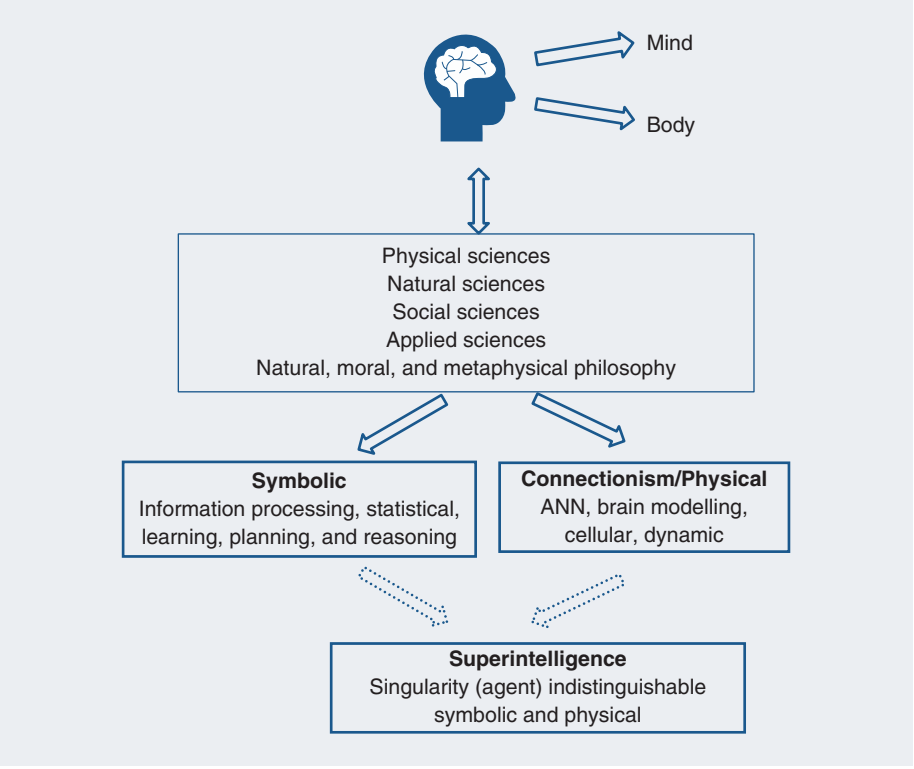
Environmental signals: derivatives from electronic, mechanical, and chemical sources.

Computational elements: symbolic and mathematical algorithms to implement the rules and processes used in sound thinking, reasoning, and behavior.

Conclusions: predicates resulting from the algorithms within the computational elements.

Actuators: signals compatible with use by external agents, as well as electronic, mechanical, and chemical mechanisms.

Figure 12.4 Approaches and goal of intelligent agents.



Machine implementation of human cognitive abilities is an ambitious and challenging objective of AI. Its ultimate goal includes nothing less than total integration, understanding, and representation of animal behaviors, namely, thinking, feeling, speaking, symbolic processing, remembering, learning, knowing, consciousness, problem solving, planning, decision making, and even the unconscious (brain). As noted in Figure 12.4, alternative approaches to understanding humans are possible.

PRACTICAL AI

To understand AI, we describe a number of fundamental topics that mirror human activities:

- Devices (hardware): comparable to the brain in a human as described throughout this text
- Information: facts also used by humans

- Logic and reasoning modes in humans
- Algorithms: processes or sets of rules for problem solving. This is akin to thinking, problem solving, and decision making in humans.

An entity, resembling humans, in which to embed some or all of these elements—a robot—is explored in Chapter 13.

Introduction

A systematic characterization of IAs that clarifies their key features is found in Table 12.2. In general, computer programs that are identified as AI often relate to prediction (e.g., based on correlations within large data sets). The range of such programs and uses has expanded rapidly since their origins in the middle of the 20th century: financial trading systems, languages and their uses, scheduling systems, manufacturing, commerce, medicine, social systems, and image processing, to name just a few. Program outputs can be used as inputs to other programs or machines (e.g., robots) that execute a desired task. A number of the cognitive and neurological aspects of the elements in the table have been explored in earlier chapters (e.g., artificial neural nets [ANNs] as well as their operations; back backpropagation discussed in Chapter 7; "expert systems" in the form of problem solving [in Chapter 5]). In this chapter, topics will relate to the architecture and algorithms that imitate human cognitive processes and behavior.

Table 12.2 Intelligent agent characterizations.

Broad Topic	Subtopic	Resolution
Definition	Human characteristics	Psychological skills: perception, association, prediction, planning motor controls
	Uses	The "Internet of things": IoT (connections to physical sensors, environment)
	Aims	Technological (useful purposes sometimes employing methods unlike human minds) Scientific: helping answer questions about human beings, other living things Neuroscience: theories of mind–brain; how the physical brain works; what the brain is doing; what computational questions the brain is answering; sorts of information processing that enable it to function Biology: artificial life (computer models of differing aspects of living organisms); bodily form entablements Philosophy: mind theories (mind–body problems); consciousness; machine intelligence
	Virtual machines	Information-processing system (the human mind—the set of mutually interacting virtual machines running in parallel, developed or learned at different times implemented in brain)

(Continued)

Table 12.2 (Continued)

Broad Topic	Subtopic	Resolution
Types of agent[a]	Classical	Also called symbolic, good old-fashioned AI (GOFAI); classical + statistics can model learning, planning, and reasoning
	Connectionism	ANNs (hybrid: a theory of human action may be viewed as a continuous switching between connectionism and symbolic). Neural nets are helpful for modeling aspects of brain, doing pattern recognition, and learning.
	Evolutionary programming	Emulates biological evolution and brain developments. Programs that can change (improve) themselves.
	Cellular automata	Models development in living organisms; closer to biological systems than psychological systems
	Dynamical systems	Models development in living organisms; closer to nonreflective behavior than to deliberative thought

[a]Can the singularity embrace all of these modalities? Progress in general-purpose AI (artificial general intelligence or AGI) requires deep understanding of the computational architecture of minds. The term *strong AI* has been used to describe artificial intelligence where the machine's intellectual capability is functionally equal to that of a human (e.g., AGI).

The rate at which developments have proceeded particularly in the 20th and 21st centuries has increased exponentially over time—a phenomenon noted in the writings of the futurist Ray Kurzweil (and others) (Kurzweil, 2012), who anticipates the eventual development of a machine that will exceed the cognitive capabilities of the human brain. These developments are strongly influenced by the societal needs associated with war and commerce, as reflected in accelerated advancements in technology.

Various cognitive models (Chapter 3) explore the relationships or distinctions between human consciousness and the brain. The ultimate integration of the physical and cognitive facets of humans within an artificial medium provides the basis for an "intelligent agent." Such an embodiment in "robotic" (behavioral) form is explored in Chapter 13. For an IA to truly reflect intelligence, philosophers argue that it must account for human consciousness. Three major theories (identity theory, behaviorism, and functionalism) are discussed at length in Chapter 3.

These processes comprise a broad cognitive and behavioral spectrum of living systems. Ultimate achievement of this lofty mission is referred to as "strong AI" and is summarized in the seminal work, *Artificial Intelligence: A Modern Approach* (Russell & Norvig, 2020):

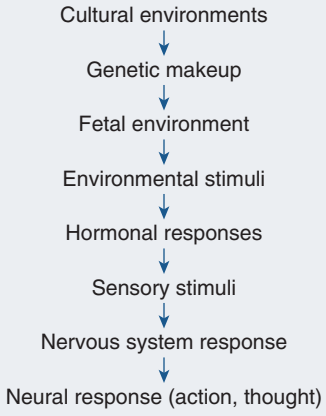
> Once we have a complete, comprehensive theory of mind, it becomes possible to express the model in machine . . . (program) . . . form.

However, Robert Sapolsky (2017) admonishes us in his seminal work, *Behave:*

> When you explain a behavior with one defined discipline, you are implicitly invoking all the disciplines—any given type of explanation is the end product of the influences that preceded it.

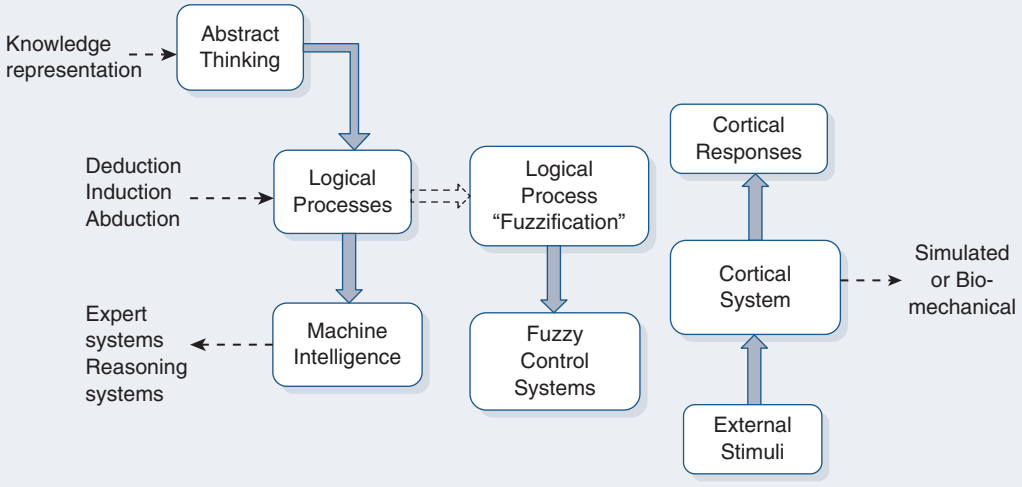
Understanding a particular behavior cannot be studied in isolation. Figure 12.5 details the complexities. This topic is explored in Chapter 1.

Figure 12.5 Behavioral complexities after Sapolsky.



AI Implementation ("Hardware")

Figure 12.6 Fundamental AI system design.



Historically, expert systems constituted the principal design of an "AI" system and were designed to emulate the decision making of human experts. Such machines relied on bodies of knowledge that were composed of facts, observations, and skills (procedural). Such machines accepted information from numerous sources: perception, memory, education, and practice. And combining these facts with logical processes that included deduction, induction, and abduction (all discussed below) could make decisions or offer advice (e.g., medical, human actions) to a user that is generated using the elements of the "machine intelligence" shown in Figure 12.6.

When there is uncertainty with regard to the information, a modified logical processing process, "fuzzy logic," can achieve the control envisioned by the cybernetic machines anticipated by Wiener, including advice or actions analogous to that of an "expert system."

Figure 12.6 depicts fundamental IA architectures, the hardware or machinery of AI. The diagram describes two distinct approaches to AI: one that proceeds from abstraction of cognitive processes (shown with arrows descending) and is referred to as the "top-down" model. Such top-down designs—whether hardware based or implemented in software (programs)—rely on a structure with two elements: facts and rules that act upon such facts. Facts rely on information, and rules reflect logic. A simple example from everyday activities provides an insight:

Fact (given): I see a red traffic light.

Rule: IF (I see a red traffic light) THEN (stop the car I am driving at the light).

The key element of the rule contains an IF–THEN phrase. The IF–THEN structure is known as a production rule (see Chapter 1). Each element includes an argument or predicate. (Predicates are "phrases" or statements whose truth or falsity can be determined.) If we can abstract human behavior to form a framework of IF–THEN rules, then coupled with observable facts, we may build a machine to automate such behaviors. Alan Turing introduced this concept in 1956. It supports drawing conclusions, proving theorems, or replicating the human activity. While the example noted above seems particularly simple, the predicates and phrases may themselves include additional combinatorial expressions using such logical operations as AND, OR, and NOT. This provides a basis for far more complex architectures and expert recommendations.

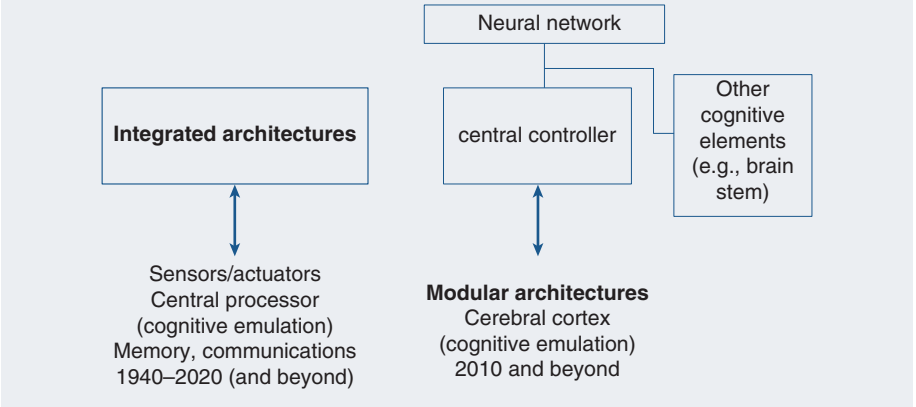
For example, the top-down architecture employs successive decomposition. It is characterized by "information processing." (Information is further discussed below.) Start with the problem to be solved and follow logical sequences of finer details implied by the concept. (This was described in Chapter 5.) A sequence for a trip might be arranged for housing, travel method, securing a ticket, and completing transport. These form the successive decompositions of the original trip.

Alternatively, an exact replica of a human cortical system with its key role in attention, perception, awareness, thought, memory, language, and even consciousness can be used to emulate intelligent responses.

A second approach—bottom-up—is closer to an "organic" representation of human cognitive processes. Bottom-up arrangements employ integrating a series of elemental or partial solutions resulting in the resolution of the original problem. (This is one approach used in deep learning algorithms subsequentially described.) Some refer to the bottom-up approach to consist of "building a brain" and letting the resulting machine learn to solve problems as a human would. For example, information enters our eye and the brain builds (recognizes) an image that then may be further processed to create a perception. In fact, because of the human "blind spot" in the visual system, the brain "makes up" the images that we "see." With respect to cognitive science, the goal of both methods is to be able to "reproduce" animal behaviors. If successful, the entity could be indistinguishable from human performance, and any "intelligence" that it exhibits could "mirror" animal intelligence.

The physical platforms on which IAs are executed have changed radically during the past 20 years. Figure 12.7 summarizes one aspect of the evolution. Since the mid-1940s, platform capacities have greatly increased. Their architectures remained relatively consistent, with central processing divided into integrated elements and graphical processor units (GPUs) assigned tasks such as rendering or creating images, animations, and graphics for display on the system's viewing screen. More recent developments now include distinct modular computational elements that reproduce other cognitive features, ANNs in particular. Development of "supercomputers" and emerging quantum computing have greatly increased computer system capacities—the amount of information processed (Shannon bits)—with greatly reduced computational times. Some analyses now consider the capacity of such machines to approach that of humans (Patterson & Hennessy, 2012).

Figure 12.7 Intelligent agent platforms.



Information and Intelligent Agents

IAs rely on "intelligence," and "information" is a key feature of intelligence. With such information, an AI can be taught—or learn on its own—to identify and imitate human behaviors. The objective of an IA is to reproduce human cognitive and behavioral capabilities in machine form. In relation to this goal, several characteristics of human cognitive ability must be considered, namely, the nature of knowledge, intelligence, information, and data. These concepts are summarized below:

- Knowledge: facts, skills acquired through experience, or practical understanding of a subject; understandings; recognition; perception
- Intelligence: ability to acquire and apply knowledge and skills
- Data: facts and statistics collected for reference or analysis or communicated to another entity

With respect to IAs, information is a quantitative attribute of knowledge and intelligence. Fundamental consideration of information measurement was introduced by Claude Shannon (1948)—including its associated attribute entropy—which conveys a measure of a variable (statistic) that exhibits "surprise" or "uncertainty." (Note that Zadeh's methodology [below] includes a measure of confidence in a variable.) Shannon explored these topics in relation to the importance of being able to communicate such knowledge or use it in calculations, and he introduced the "bit" as the unit of uncertainty. A simple example explains these ideas. We wish to convey the outcomes of tossing a coin. (It could be a single toss or several tosses.) Our coin may be biased (weighted) such that it will always come out as heads, or it may be biased so that it always comes out tails. Such a coin would ruin a game that resulted in such biased outcomes. We describe the bias by a probability of the outcome. (A fair coin has an equal probability of landing on the head or the tail.) In this example, we assign a probability of "p" to landing on heads. The only other outcome is a tail, and its probability is therefore $(1 - p)$. For example, if $p = .5$, then $(1 - p)$ is .5, and there is a 50/50 chance of a head or a tail; this defines "fair coin." The extreme cases—always heads or always tails—have an entropy of 1 bit because there is no "surprise" (uncertainty). Shannon defined p as a measure of the amount of entropy (or its information content). In this example, when $p = .5$, the toss will have the most uncertainty and will contain the most information and reflects the most uncertain outcome. Experiments have the most information when all outcomes have equal probability because all outcomes are equally likely and uncertainty is maximized.[4]

One important use of this concept relates to determining the capabilities or capacity of an AI system (as well as that of a human). George Miller (1956)—a Harvard cognitive psychologist—noted that a human can retain between five and seven "objects" in short-term memory. He referred to this as the "magic number 7 ± 2" based on his experiments.

[4]Shannon's mathematical formula can be found at en.wiktionary.org/wiki/Shannon_entropy.

It is called "Miller's law." If a human must retain these (seven) objects—say numbers—for 10 seconds, then the individual would need to have an information-processing capacity of (7/10) or 0.7 bits of information per second that it can remember. Such capacity requirement also applies to a computer or system where machines may need to be able to process as many as trillions of bits of information per second. (A factoid: The piano is the ultimate instrument in terms human capacity.)

Logic, Classical and Fuzzy

Table 12.3 summarizes three fundamental operations available for logical processes. The table includes two columns; one reflects classical logical operations, and a second mirrors the important contribution made by Lotfi Zadeh (fuzzy logic). Treatment of the "truth" (certainty/uncertainty) of a predicate is attributed primarily to Lotfi Zadeh (1965). He described such logic as "fuzzy logic" to reflect the fundamental uncertainty of events. (When diagnosing an illness, a physician may not consider a patient's body temperature of 99°F to constitute true "febrility" even though 99 is above what is considered normal [98.6].) An example of Zadeh's fuzzy logic will be considered below, particularly as it relates to reasoning. In Zadeh's formulation of the logical operation, the result of the operation reflects the belief that "best" satisfies the predicate operation.

Reasoning Modalities

The architectures of Figures 12.6 and 12.7 subsume human reasoning, which is not inherently "simple, neat, or impeccable."[5] Such platforms support three distinct logical methods for drawing conclusions: deductive, inductive, and abductive. These are summarized in Table 12.4.

Table 12.3 Summary of logical operations.

Logical Operation	Predicate Operation	Traditional Logical Result	Zadeh's Fuzzy Logic
AND	A AND B	TRUE, only if both A and B are TRUE	Choose the lowest belief, A or B
OR	A OR B	TRUE, if either A OR B OR both are TRUE	Choose the highest belief, A or B
NOT	NOT A	Inverts the status of A; if A is TRUE, result is FALSE; if A is FALSE, result is TRUE.	Choose a value that reflects the certainty of (1 – belief of A)

[5] https://evolvingthoughts.net/2010/10/27/human-reasoning

Table 12.4 Logical reasoning.

Logical Method	Definition	Example	Limitation
Deduction[a]	Reasoning from a series of logical statements to reach a logically certain conclusion. If (A [predecessor]) is True), Then (B [follows] is True).	All men are mortal; Socrates is a man; therefore, Socrates is mortal. (A classic example)	In AI, it may fail to include all relative facts and therefore draw the wrong conclusion (e.g., the self-driving car was not aware of an object).
Inductive	A logical conclusion follows from several premises that are found to be true all or most of the time.	All birds that I have seen have wings and fly. Penguins have wings but do not fly. They are not birds.	Penguins have wings but do not fly. They are "flightless" birds but are birds nevertheless.
Abductive[b]	Starting with a series of observations (and facts), a conclusion follows that finds the simplest and most likely explanation.	Such logic was widely used in expert systems. MYCIN would provide a list of diagnoses together with the confidence in the diagnosis.	It can also lead to errors. In a legal case, the jury hears all the evidence and draws what its members think is the most likely conclusion. This could be in error, and the same evidence might have been the result of a different precipitating event.

[a]Considered to be perfect logic: You cannot prove a true predicate false and you cannot prove a false predicate true.

[b]Much of human decision making is abductive.

THE LEGACIES OF TURING AND ZADEH

Turing's State Variable Approach

Alan Turing is remembered for the "Turing Test." It was an attempt to fool a judge and not to create a conscious human equivalent such as one model that was proposed by Joseph Weizenbaum (https://en.wikipedia.org/wiki/ELIZA).

However, Turing's true legacy is much more relevant to intelligent agents. He suggested the possibility of a "universal computing machine" (Turing 1937), a seminal principal on which the design and function of numerous computer-based automata are built. The universal computing machine is characterized as a machine that can "simulate" any computer algorithm no matter its complexity. If one can formulate a method (algorithm) for solving a problem, then it is possible to build a machine to implement the solution. Its application in the cognitive sciences as well as computing and engineering technology

of the past century has been profound. A contemporary representation of such "state machines" is depicted in Figure 12.8. It is vital for the design and analysis of physical machines.

A series of states reflect conditions within an environment. Signals from the environment determine transitions between states. In turn, signals from the various states control the environment. An example explains these interactions and is summarized in Table 12.5. Consider an automatic dispensing system; the following sequence of events defines how the system expects to operate. (It is simplified as a number of events are not considered, such as user mistakes, change to be returned, or machine operational failures, among other considerations.) Sensors are needed to detect user actions, and effectors (mechanisms) are needed to accept and evaluate coins, as well as dispense items.

Figure 12.8 Contemporary Turing computing system.

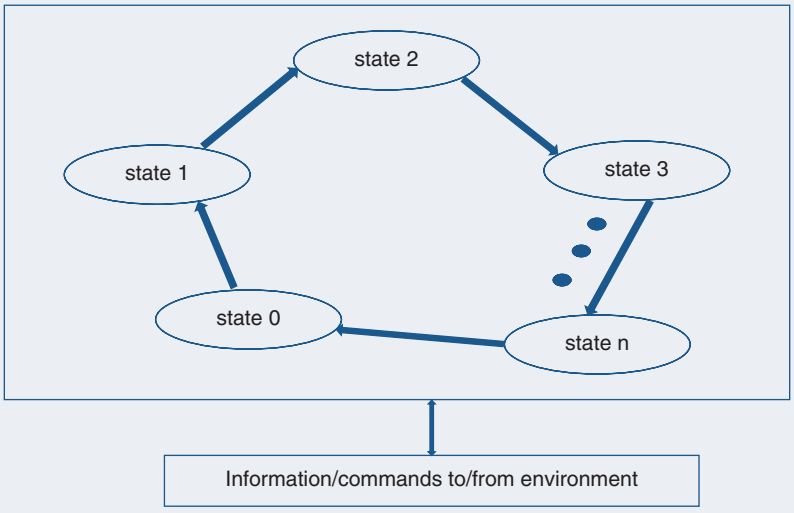


Table 12.5 Logical design of dispensing system.

State	Condition	Command	Action	Next State
State 0	Awaiting use (reset)	Insert coin	Insert coin	S1
State 1	Insert coin	Accept coins	Make choice	S2
State 2	Make choice	Choose item	Check amount	S3
State 3	Dispense item	Dispense item	Item received	S4
State 4	Item received	Item received	Reset system	S0

Within the system, a series of IF–THEN commands are executed to implement the constraints depicted in the diagram and derived directly from the table. As an example: If the machine is in state 0, the following rule applies:

IF (awaiting use AND user deposit coin) THEN (proceed to State 1 [coin insertion state]).

Similar IF–THEN rules can be extracted from the state table to complete the design. The finite-state diagrams (such as the one noted in Figure 12.8) can be reliably and easily transformed into a program of executable instructions that, when stored in a general-purpose digital computer with associated electrical and mechanical elements, would accomplish the desired operation. (For completeness, a designer would want to provide for malfunctions such as failure to recognize exact amount of money deposited or other malfunctions in the machine.) The potential to eliminate such jobs as toll taker is obvious and should not be minimized.

If human behavior and/or thinking could be represented in such a way—that is, a series of "IF–THEN" activities characterized by sensing the environment and either making decisions or generating actions—then there is reason to believe that a machine could be built to replicate (or replace) the animal behavior—"perhaps" even the neocortex or the entirety of the brain. The system described above operates in strictly serial fashion, that is, one operation at a time. However, the human cognitive facilities as embodied in the neocortex function in a "parallel" manner (as discussed later in this chapter as well as in Chapter 7). Parallel activity means that recognition of states (patterns) and consequential actions occur concurrently ("in parallel")—such capability produces overall faster results. The limitation of (slow) serial operations is one argument used to maintain that such machines will never "supplant" or even emulate the human brain in any practical way (because of the number of interconnections in the brain).

It should be noted that a basic cognitive model of an IA—a mechanism that incorporates intelligence and complements the finite-state paradigm—was proposed by Kenneth Craik (1943), a philosopher and psychologist who developed such a cognitive model before Turing's paper.

The practical (behavioral) approach to AI that underscores Turing's universal computing concept is based on the principles of logical reasoning. These systems subsume a means of knowledge representation that is compatible with the digital base of the computer, supportable facts, and a system of rules that permit these IAs to plan, make decisions, and act upon the results (e.g., behavioral actions) that derive from such models. Knowledge, facts, and rules must be free of referential ambiguity. (Additional discussion of representation and meaning, including referential ambiguity, is addressed in Chapter 1 in additional detail; in particular, it focuses on connections between the external and cognitive worlds.)

A complete study of knowledge representation should include extended considerations. One such important formulation is discussed in Chapter 7. Other formulations include frames, cases, scripts, and others. Each of these models is fashioned because of the application requirements. Descriptions are needed to address correspondence between the surrogate (computer entity) and its equivalence in the real world in a way that is compatible with the computer algorithm.

Fuzzy Problem Solving

Fuzzy logic addresses one important element of data, namely, "what happens when we do not have complete confidence of the truth of the facts in the rule base?" Uncertainty is a feature of human perception and needs to be considered when making judgments and/or decisions. This is a phenomenon that motivated the development of fuzzy logic as well as the need to incorporate some of its implications in expert systems (ES). Issues related to uncertainty have been studied by Amos Tversky and Daniel Kahneman (1974). Uncertainty appears when not enough definitive information is available and when decisions are required for extremely complex problems (e.g., too much information is available). In such circumstances, humans employ "heuristics" to make decisions. (A heuristic is a speculative formulation guiding investigation or solution of a problem.) Heuristics may lead to biases, and although they can be incorrect and/or harmless, they may lead to errors of judgment (e.g., prejudices when interacting with other people). Within an ES, heuristics may be employed through "probabilistic" representations. The analysis proposed by Lotfi Zadeh includes heuristics determined by systems of belief ("membership functions"). Discussion of this human characteristic is provided in the chapter on evolution.

The "belief" in a variable (e.g., thought or action) was formally introduced by Lotfi Zadeh (1965). The resultant logic is characterized by "approximate reasoning" and is currently the basis for enumerable "smart" systems found in daily use (e.g., automated control of transportation and infrastructure systems, medical [robotic surgery, anesthesia], cell phones).

Zadeh's concept is referred to as the "theory of fuzzy sets" because it involves logical operations that reflect "fuzzy" thinking as opposed to "absolute" truth or falsity of predicates—what does "going too fast" mean for the speed at which a human is running? Zadeh introduced the idea of a membership function that reflects the extent of the truth of a predicate variable; it has a range from 0 to 1, where "0" implies absolute falsity (nonbelief) and "1" indicates total belief ("absolute truth"). Subsequently, he proposed modifications of operations for logical calculus. The modified operations are shown in Table 12.3.

A simple example demonstrates how it is employed. Consider the choices a person makes when deciding which employment opportunity to select. An individual may consider three—possibly more—factors when choosing which employment opportunity to pursue. In this example, salary, interest, and convenience are job characteristics of importance. The individual rates these factors for a job where "interest" involves the nature of the assignment, and "convenience" may include travel time, location, and other factors. (We restrict the ratings to values between 0 [low] and 1 [high], where 0 is the lowest judgment rating and 1 is the highest judgment rating.) An individual is presented with three employment opportunities with the ratings shown in Table 12.6 and has made determinations regarding the "importance" of the various parameters.

We can calculate the overall belief in the quality of each opportunity. The quality of a job satisfies the logical AND operation (i.e., the qualities of the salary AND interest AND convenience).

$$\text{Quality of a job} = (\text{salary}) \text{ AND } (\text{interest}) \text{ AND } (\text{convenience})$$

Table 12.6 Employment membership beliefs.

Job Number	Salary	Interest	Convenience
1	0.4	0.8	0.5
2	0.7	0.5	0.2
3	0.5	0.6	0.3

Note: Membership value is a judgment about each parameter. For example, it might be the result of statistical surveys of the parameter.

The "fuzzy" AND operation consists of taking the smallest belief or weight among all variables, providing us with a "conservative" choice. (See the "Fuzzy" AND operation in Table 12.3.) The results are therefore:

Job 1 quality = 0.4

Job 2 quality = 0.2

Job 3 quality = 0.3

The final selection would be to choose the job that produces the best overall quality. This translates into the following fuzzy operation:

Choice of best job = (quality of Job 1 OR quality of Job 2 OR quality of Job 3).

Following the rules set forth by Zadeh (Table 12.3), the result of an OR operation is the variable with the highest belief. In this case it would be Job 1—even though it includes the worst salary that many individuals might normally reject without consideration of the other factors. Zadeh's contribution to AI has significantly advanced practical development and implementation of IAs, as history has demonstrated.

INTELLIGENT AGENTS THAT THINK, LEARN, AND MAKE DECISIONS

The development of IAs follows a symbiotic path; greater computer capacities lead to more ambitious attempts to re-create human behaviors, which in turn make new demands on computer capacities. The architectures (Figures 12.6 and 12.7) reflect such evolution. Central to emulating behavior requires IAs to reproduce the unique ability of human learning. Keep in mind that humans are capable of mistakes; it behooves us to consider that computers may make mistakes. This has been demonstrated by such occurrences as plane crashes due to algorithmic error. Currently, machines can imitate a great many human attributes (e.g., read, write, smell, talk, see, touch, move,

understand emotions, make choices). The state of their development now permits them to "compete" with humans in learning new material, playing games (e.g., chess, GO, video games), and even thinking. However, machines are not considered to exhibit "consciousness" in part because understanding of this animal phenomenon is not totally resolved in science. Using a machine to facilitate learning falls into two major categories. These are summarized in Table 12.7.

Table 12.7 Types of learning.

Type	Features
Machine learning	Human designated data features are learned (e.g., bananas are yellow speckled with brown, curved body); new data are identified based on learned features.
Deep learning (neural networks)	"Top-down" methodology: Objects are identified (e.g., cat images are identified in many contexts); new samples are subsequently recognized from learned abstractions. "Bottom-up" methodology: Simple features are learned (image line, edge); additional layers of greater abstraction features based on simple features (cube as combinations of lines and edges); new data are recognized from these abstract objects.

FUNDAMENTAL CONCEPTS OF THE IA

The idea that a computer could learn to play a game as well as a human was suggested by Arthur Samuel (1959) in his groundbreaking work at IBM, where he demonstrated that one could "teach" a computer to play a game (checkers). He proposed that the tactics he used could be generalized. He employed a search tree[6] (similar to the solution described in Chapter 5). However, because the computers then had limited memory, he needed to enhance the simple tree methods to avoid "computational explosion" (i.e., a prohibitive number of calculations that exceed the computer's capacity as well as the time it would take to follow all alternatives). He stated that his work "[could] verify the fact that a computer can be programmed so that it will learn to play a better game of checkers than can be played by the person who wrote the program." This remains a goal of machine learning. The machine "scored" the current board position and eliminated those paths with the worst score for further evaluation. (This became known as "pruning" and is still used in what is called "alpha-beta" pruning.) He subsequently explored other methods, including a unique approach of having the computer play games against itself to improve

[6]A tree generally consists of a flowchart of nodes consisting of events or attributes of a process or model together with the paths that join these elements.

its performance. It is an example of having a machine learn without additional help from a human. Elements of machine learning include using statistical modeling to detect patterns, improving outcomes based on data, and incorporating empirical information. Such techniques are central to "data science." Humans are still very much involved in these elements because they choose the data that the computer uses in the learning phase. If the data are flawed or biased, then results will not be satisfactory. The overall learning process includes the following:

- Expose the machine to data associated with the intelligence to be emulated or simulated and generally chosen by the designer. (Simulation comprises a computer model for the purpose of study, and emulation attempts to match or surpass a behavior.)
- Select an appropriate learning algorithm.
- Set parameters associated with the algorithm (e.g., target score).
- Machine adjusts judgments based on a system of trial and error.
- Machine forms future decisions based on the experience gained during learning, which mimics human decision making.

This is similar to the way we train animals (Miltenberger, 2008). College admission officers seeking advice on candidates using computer "intelligence" might train a machine using SAT, grades, recommendations, cultural backgrounds, experience, extracurricular activities, and other factors ("legacy"). They will use such information to estimate an applicant's success in obtaining a degree from their institution (or being successful upon graduation from the institution). Machine learning supports AI in several ways: perception models, natural language processing (NLP), and data mining (a subject related to searching for patterns in large databases).

There are three basic techniques for computer learning (Mohri, Rostamizadeh, & Talwalkar, 2018):

- Supervised:
  - Provide test data (samples) with features that characterize the intelligence to be learned.
  - The computer forms a model. If outcomes are not satisfactory, a new training set is introduced to refine the outcomes. This is repeated either a fixed number of iterations or until the results are "satisfactory." (Note: They do not have to be perfect. A very skilled athlete may not always have a perfect performance.)

    Typical algorithms for the learning process include: trial and error, linear regression, Bayesian analysis, decision trees (see Chapter 5), and neural networks (NNs) (see Chapter 7).

    Variations of such models are also possible.

- Unsupervised: No data classifications are provided; the computer must discover the hidden patterns on its own. (While this may seem extraordinary, we present an example of how the computer may achieve proper outcomes.)
- Semisupervised: Several well-labeled examples are provided to "self-train" the system; the result is "partially trained"; the system is exposed to additional "unlabeled" data; the resultant combination of self-learned, and the added task is called "pseudo-labeled"; and the resultant model is one that combines supervised and unsupervised learning.

  In all algorithms, a critical reinforcement cycle improves outcomes.

The agent's results are evaluated by giving it a score; future tasks may be more difficult, and if the results fail, the goal could be reduced; a new strategy attempts to increase the score; this may continue until some optimal strategy results. As an example, consider a self-driving car that may be given a score for a given task but fails the task; the car must seek a strategy from the data that will pass the task in the future; repetition of the task may then be successful. (This is akin to "instrumental conditioning," a concept from behavioral psychology [Schacter, Gilbert, Wegner, & Hood, 2011].)

Basic Models

Computer-based IAs use models based on NN technology. This was explored in Chapter 7 in detail. A conceptualized diagram is shown in Figure 12.9.

From this model, two simulation modalities have evolved. Both involve the neuronal weights shown in the computational element.

- Learning machine: The weights of the computational element are adjusted as learning progresses.
- Deep learning (DL): The organization of the weights of the computational element is adjusted as learning progresses. Since the number of nodes and their interconnections define the "algorithm" performed by the machine, changing these elements is equivalent to changing the model. Learning progresses by finding. In effect, the computer is developing the algorithm.

The characteristics of machine learning and DL machines are summarized in Table 12.8.

Figure 12.9 Conceptual model of an IA computational element.

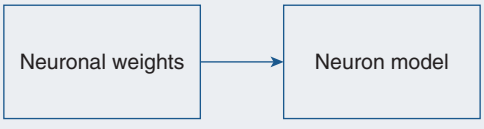


Table 12.8 Characteristics of deep learning and machine learning.

Parameter	Deep Learning	Machine Learning
Data	Uses large amounts of data	Learns on fewer samples
Accuracy	High (lower error rate on untested data)	Less accurate (greater error rate on untested data)
Time to train	Longer	Shorter
Computational needs	Requires larger machines (e.g., requires ability to display graphic information)	Can operate on lower-cost computer systems
Limitations	Lacks common intelligence, general intelligence, domain information	Needs large amounts of carefully crafted data; learning generally supervised Poor transfer learning Opaque Poor performances on symbolic reasoning or planning

The goal and promise of DL machines seek to reproduce "thinking" that is used to solve a problem or identify an object (e.g., How did the machine beat the expert GO player?). DL, however, has critics. Deep learning may result in very "shallow learning." There is extraordinarily little theory underlying it. The goal of any of these systems after "learning" is the following: Apply an input to be identified; the system performs an "arithmetic" transformation to produce a (hopefully) accurate result. To date, researchers cannot describe what the machine does in "functional terms"—a mathematical equation. There is considerable research in trying to develop such mathematics but, to date, has not met with success. That is why one can follow the machine's "thinking" in GO, but it cannot determine the mathematical formula—the "why" of the thinking. For such machines, there are no guarantees of convergence (since we are, after all, talking about nonlinear optimization in high-dimensional spaces) and no performance guarantees of any kind (say, compared to what you get with other areas of machine learning, like kernel methods, sparse linear models, etc.). The DL critic Vladimir Vapnik[7] notes the following:

> The most practical thing in the world is good theory", and that's currently not available in deep learning. If deep learning is the best solution the machine learning community can do, as a card-carrying member of this research community for over 30+ years, I'd have to say we are in serious trouble!

[7]This quote can be found by an Internet search for "How will humanity be impacted by deep learning . . ." at https://www.quora.com.

Machine Learning and the Data Sciences

Matthew Stewart[8] notes,

> Machine learning systems are becoming increasingly employed in complex high-stakes settings such as medicine (e.g., radiology, drug development), financial technology (e.g., stock price prediction, digital financial advisor), and even in law (e.g., case summarization, litigation prediction).

DL is a new and emerging methodology. Machines that use DL depend on and are often limited by the capacities of the system. As an example of the capacities of emerging systems, consider Microsoft's introduction of the largest DL NN system currently in existence—the "Turing-NLG" —in February 2020 (Rosset, 2020). Its operations support the equivalent of 17 billion neurons or about one fifth of the number in the human brain. (This does not represent the 100 trillion connections in the brain.) A NN that supports 17 billion neurons will enable such machines to study very extended and complex databases. Too much data can, in fact, overwhelm a machine's ability to interpret information. Concentration on data in one circumstance might cause errors when presented in another setting (e.g., face recognition in light environments vs. recognition of the same face in a dark background). We limit explanation of data interpretation to relatively simple cases.

The Trial-and-Error Learning Method

One of the simplest forms of problem-solving methods is trial and error. When faced with a problem to be solved, an agent—including animals, humans, or (intelligent) machines—successively tries various responses until some success is achieved. These responses are seemingly random in nature (i.e., trial and error); however, each successful choice—getting closer to the objective—is strengthened until the final trial sequence is successful in achieving the goal. A classic example of such learning is maze learning. Starting from a designated point in the maze, the agent is required to reach a goal in the most direct way with minimum errors. The trial-and-error process requires elimination of paths that do not lead to the goal (e.g., elimination of blind-allies). The trial-and-error algorithm follows from the stimulus–response theories of Edward Thorndike (1898), who described such behaviors as "trial-and-error accidental success." Within the past decade (*Psychology Today*, March 7, 2016), neuroscientists (University of California, Berkeley) have recorded brain images of active real-time learning in mice during a trial-and-error problem. Using advanced microscopy techniques, they were able to make time-lapse movies of the orbital prefrontal cortex that supports how a mouse learns a new strategy for finding hidden treats in a foraging task. The movies reveal the gradual resculpting that took place in the frontal lobes within the cerebrum.

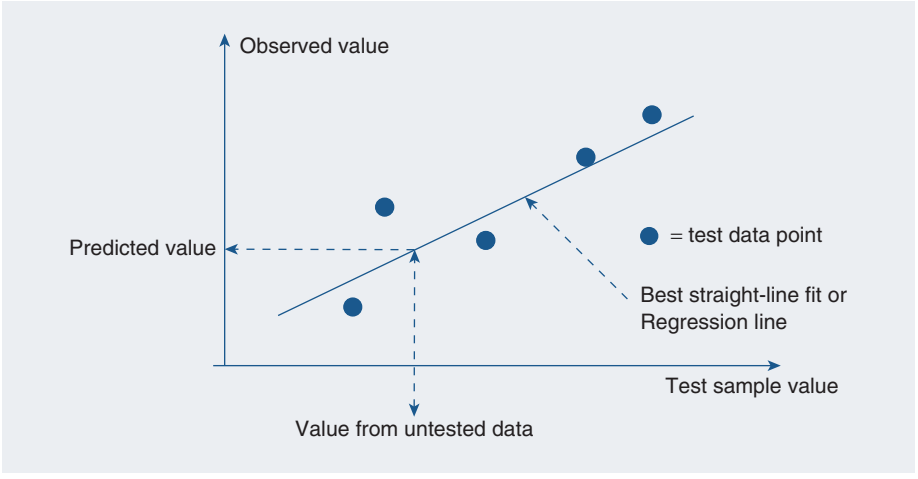
[8]https://towardsdatascience.com/guide-to-interpretable-machine-learning-d40e8a64b6cf

A Simple Algorithmic Method (Regression Line Algorithm)

During training, data from the test set may cluster around a well-defined region that can be represented as a function (or equation of a straight line). This is the case in Figure 12.10.

A regression line summarizes the relationship between the data values of the test samples and their dependent variable (the y-axis on the graph in Figure 12.10). It is a single straight line that best "fits" the data points (on the x-axis). It is that one line having the smallest combined overall distance from the line to the test data points. Such analysis finds a variety of applications in biology, behavior, environmental and social sciences, and business. Regression line analysis has a long history of study, and its properties are well understood; training an agent using this algorithm is efficient. Actionable results from such analysis are easy to initiate.

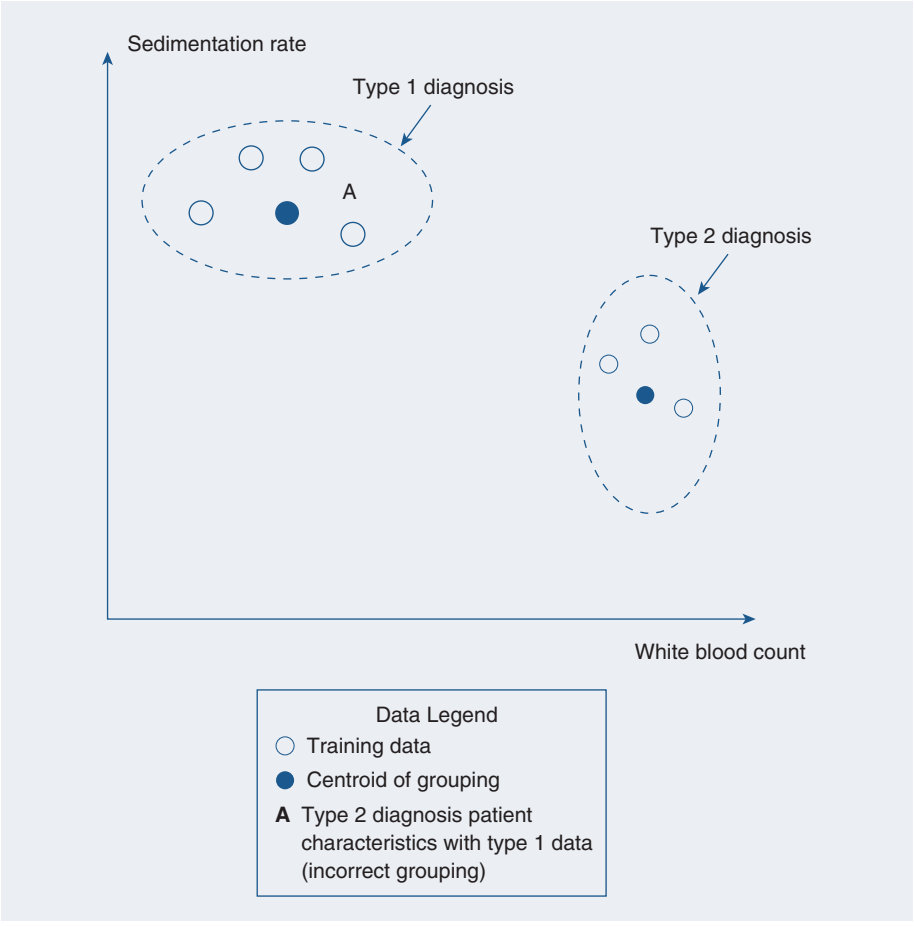
Figure 12.10 Regression line estimation of events.



Data Clustering

When data do not align themselves in the "scattered" arrangement shown in Figure 12.10, another strategy may be employed. This can be a productive strategy when employing unsupervised learning. Figure 12.11 reflects the graphical representation of the algorithm when applied to a specific diagnostic problem. The abscissa (x-axis) represents readings (number) of the white blood cells from a patient's blood, and the ordinate (y-axis) represents values of a patient's sedimentation rate (inflammation level) in the blood sample. These clusters represent symptoms from two different diagnosis (e.g., bacterial infection vs. autoimmune rheumatic disease). The AI is being trained to determine which diagnosis to make in response to presentation of similar symptoms. The centroid of a group of points is the geometric center of a plane figure. It is the

Figure 12.11 The *k*-means method for classifying data.



arithmetic mean position of all points in the figure. Think of it as the point in a cutout of the shape where it is perfectly balanced. The centroid is thus "equidistant" from the centroid of each group within the training set. When test data (as opposed to the training set) are provided, the type determined is the one closest to the centroid of the type of diagnosis. If closer to the Type 1 diagnosis (autoimmune disease) data centroid, the algorithm will indicate the corresponding diagnosis. An analogous result is obtained for symptoms close to Diagnosis 2 (bacterial infection). Notice that one sample designated as "A" is from a patient with a Type 2 diagnosis (bacterial infection), yet the blood samples correspond to a Type 1 disorder (autoimmune disease). When training an AI, some data may not fit the desired characteristics. When trained on such data, it is possible that the AI will not always produce a correct result on data not previously seen (as from a new

patient). An IA, like a human physician, may not always be correct. (Perhaps more training examples are needed with additional symptoms [diagnostic refinement].)

Other examples of unsupervised methods include association analysis, social network analysis, and descending dimension algorithms but are not discussed (Duda, Hart, & Stork, 2001). While the training set does not have to be correct, it must generate an "acceptable" number of correct answers. As new, previously unseen data are presented to the agent, the centroid closest to its coordinates will identify the information. For example, blood samples from a patient are submitted, and the agent then determines the corresponding diagnosis. (It should be noted that a system dependent on only two variables [coordinates or parameters] is not likely to be very edifying in real diagnosis, and this method will fail if data do not lend themselves to clustering as shown in Figure 12.11.)

Reconsideration of Data Sources

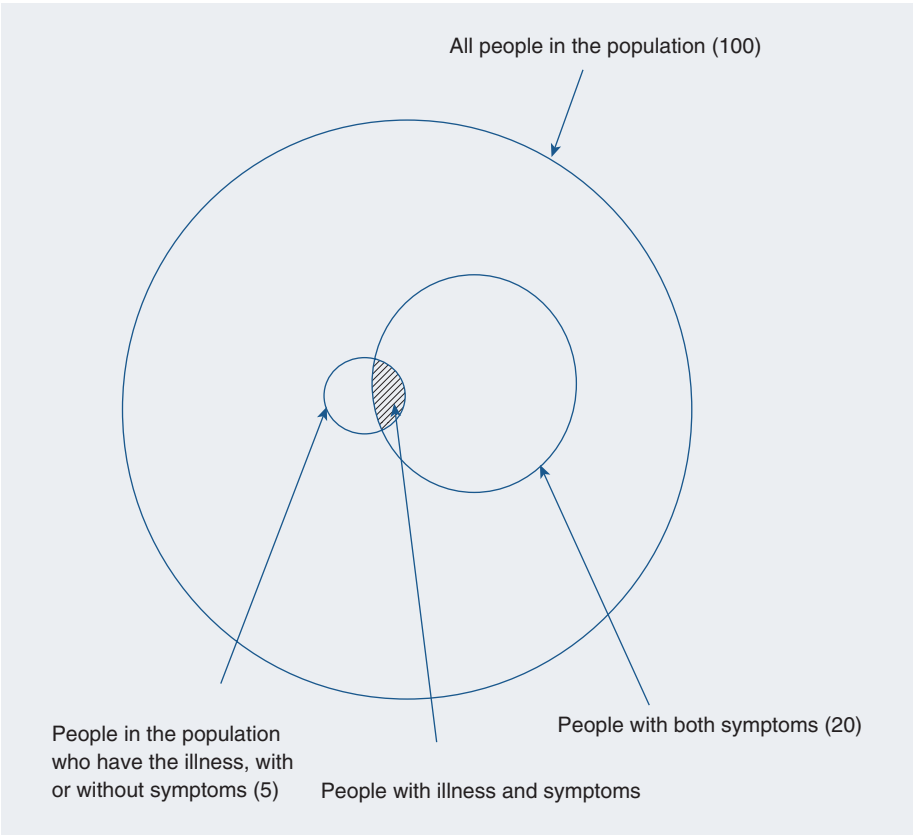
In 1763, the Reverend Thomas Bayes first questioned the significance of the evidence that leads to a conclusion (Bayes, 1763); the French mathematician Pierre-Simon Laplace formalized this conjecture. (While Laplace provided the mathematical analysis, the name carries Bayes' theorem as its identifier [Laplace, 1902].) We will demonstrate how it permits us to revise our estimates (i.e., probability) of the evidence that we used when drawing conclusions. It can help to update a belief that is based on new evidence, revise the probability of an assumption, or update a hypothesis as well. The following facts are given for drawing medical conclusions:

- Ninety percent of people with a given illness have two symptoms (e.g., fever and chills for pneumonia). When this is the case, one might be inclined to diagnose such an illness when a patient has these symptoms. This could lead us to the (abductive) conclusion that the patient should have pneumonia. But we should consider all evidence.
- Only 5% of all people get the disease in question. In other words, if we consider a population of 100 people, 5 people might get this disease even before we consider the symptoms. (You may not have the illness even if you have the symptoms.)
- People might also have such symptoms whether they have the illness or not. We discover that 20% of people could have these symptoms whether they have the illness or not. So, 20 people in our population of 100 have these symptoms, but we do not know how many of these 20 have the illness because they must have both the symptoms and the disease.

To visualize the problem, consider the diagram shown in Figure 12.12.

Bayesian analysis follows: The percent of the people who have the symptoms and the illness are calculated as follows: 90% of the people who have the symptoms have the disease, but only 5% of the population actually have the disease. Thus, 0.45 (45%) have the symptoms and the disease (0.9 × .05). But this group is only 20% of the population to begin with. Therefore, our overall confidence that our diagnosis is correct is (.9 × .05)/0.2 = .22.

Figure 12.12 Bayes' theorem diagrammed.



Our confidence in our diagnosis is only 0.22 (22%), even though we may have thought that the confidence should be .9 because 90% of people with this illness had the symptoms. This example should alert us to take many factors concerning information before drawing conclusions. These same results are found by a direct use of Bayes' formula:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}.$$

$P(A \mid B)$ = Probability of having pneumonia knowing the patient has symptoms (B)

$P(B \mid A)$ = Probability of symptoms (B) knowing that patient has pneumonia (A)

$P(B)$ = Probability that anyone has symptoms that were observed (whether or not they have pneumonia)

$P(A)$ = Probability that anyone has pneumonia (disregarding any symptoms)

In this example, the corresponding values are

$$P(B) = 0.2, P(A) = 0.05, P(B|A) = 0.9.$$

Substituting, we conclude that $P(A|B) = (0.9) * (0.05) \div 0.2 = 0.22$.

Using Bayesian analysis can be especially useful, but it must be included with other learning elements. Its limitation can be inadequate because the analysis does not, by itself, generate "novel" explanations of the purpose of an action but only fine-tunes or, at best, chooses among existing ones; "creativity" is needed.[9]

Deep Learning (DL)

With few exceptions, all animals have neurons. They can be grouped into sensory, motor, and interneurons (providing connections between neurons in the same region of the brain or spinal cord) depending on their purposes. The neuron is discussed in some detail in Chapter 7—see Figures 7.4 and 7.5 in particular. Also included in Chapter 7 is a detailed explanation of backpropagation, the methodology for adjusting the weights of the links between nodes of the neuron. Figure 7.6 defines the process whereby a three-layered neural net can "learn" to produce "humanlike" responses (e.g., recognize letters of a language). By combining a number of such layered neural nets (NNs), systems can be trained to reproduce complex human behaviors. Such networks (Ge & Wang, 2002) evolved from the control system models suggested by Norbert Wiener and advanced technologies such as solid-state computational elements. Table 12.9 summarizes key historical events regarding neurons.

Table 12.9 Significant neuronal technology events.

Year	Scientist	Significance
1887	Santigo Ramon y Cajal	Microscopic image of the neuron
1891	Heinrich Waldeyer	Introduces the term *neuron* as a fundamental element of the nervous system
1943	McCulloch/Pitts	Mathematical model of a neural network
1952	Hodgkin/Huxley	Intercellular recording of neurons
1986	Rumelhart/Hinton/ Williams	Backpropagation and use in neural networks

[9]Simpson's paradox is a phenomenon in statistics where trends appear in different data but disappears when these groups are combined. It may be resolved when causes are taken into account in the statistical model. It is often encountered in social and medical sciences (see Wikipedia).

NNs are mathematical models of the brain; they do not reflect the way in which the weights are organically adjusted. To accommodate this reality, NN architectures rely on a model originally proposed by Wiener (1961) combined with the solid-state technologies available in the late 20th century. Contemporary architectures take advantage of the greatly increased capacities of advanced computers and reflect a serial combination of the three-layered NN shown in Figure 7.5. The potential capabilities of emerging NN systems supporting DL are exemplified by the Microsoft system currently in existence—the "Turing-NLG"—described above.

DL Software

Deep learning (DL) software (Duval, 2017) is based on a form of regularized NNs (RNNs). It is one aspect of machine learning in which regularization consists of processes of adding information to solve "ill-posed" problems. Such problems are those including those in which a well-defined solution does not exist, a solution may be not be unique, or a solution is dependent on changing conditions. Regularization makes it possible for a machine to simplify models. DL systems can take advantage of hierarchical patterns in data associated with a problem and assemble ("learn") smaller and simpler patterns. Greatly truncated historical highlights of DL are noted in Table 12.10.

Each DL layer consists of an NN like the one shown in Figure 7.5. (The elements in the NN can be simulated by a set of software instructions.) That is, there are three stages within each layer (input, hidden, output). In each layer, different neurons in the layer are activated in response to the data presented at its input. In this way, each layer determines identifying elements of the input. For example, each layer might identify different highlights of a facial recognition agent. Moreover, DL machines can "discover" the highlights essential for correct recognition of a face. To train such agents, a "model" of the object—an animal that may have ears and a tail—and a set of learning data are

Table 12.10 DL highlights.

Year	Event
1965	Alexey Ivakhnenko: a working learning algorithm; supervised, feedforward, multilayer perceptron
1980	Kunihiko Fukushima; supervised, feedforward, multilayer perceptron that reduces errors in numerical recognition programs
1986	Rina Dechter: introduces the term *deep learning*; it is accepted by computer science professionals
1995	Brendan Frey: six layers trained in 2 days
2014	DL detects toxic effects of environmental chemicals in nutrients, drugs, and household products

Table 12.11 Organization of a DL system.

Layer	Feature
Input	Pixels (raw information)
First	Edges
Second	Corners and contours
Third	Object parts
Fourth (final output)	Object identity

presented to the program. If the agent makes an error, some directions are given to improve subsequent examples that will tend to correct the error. Within each layer, the neurons may respond to either true or false choices or to specific numerical values of the data that they process. The number of such (three-segment) layers connected in a serial manner now approaches 100, which greatly enhances the uses of DL systems.

For example, a DL system can be presented with several examples of an object (e.g., an animal, face), and the agent is required to learn to recognize these objects (after training) in a variety of contexts (e.g., pick out a particular face in a crowd of people). Such models are considered to have "top-down" architectures.

Alternatively, the DL system may be trained to "build" an object from fundamental elements as opposed to complete objects. This is a "bottom-up" model. Table 12.11 provides an instance of such an entity. It can be trained to extract features from raw information. In the example, four layers (three-stage NNs in each layer) are employed to incorporate detailed information into highly abstract ideas.

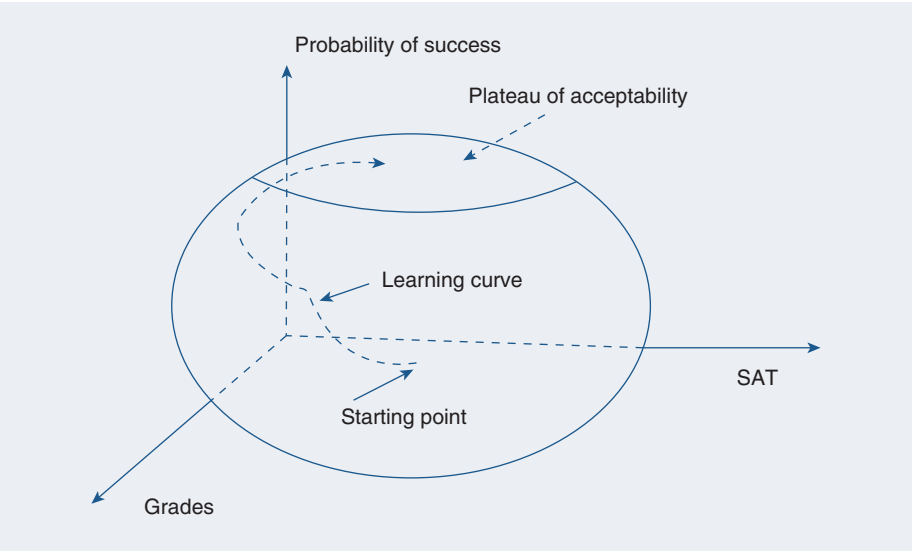
The training algorithms themselves are characterized as supervised, unsupervised, or semisupervised as previously described. The programming details are excluded.

LEARNING EXPERIENCES

The goal of machine learning is to implement a strategy to find the best (optimal) response to factual and environmental data presented to the IA. The process by which this is carried out is basically to assign a measure (e.g., "score" such as number of correct answers) to each iteration of the test set and improve this score until some "acceptable" limit is achieved; there are a number of such schemes, and a simple one will be explored here, as shown in Figure 12.13.

A simple ("top-down") DL system learns to identify the best chance that a student may have to complete an advanced degree based on two parameters: the student's SAT score and her or his grades. (Care must be taken to avoid undetected bias such as ethnic prejudice.) A data set including a great many samples of the SAT and grades as well as whether the student completed an advanced degree is presented to the

Figure 12.13 Simple training experience.



DL system. Using documented experience, the system can find the probability of completing degree requirements based on the parametric values. (Each test case also includes the student's success or failure.) This is compared to the current "guess" that the system has established. Based on the error, the NN weights are adjusted so that the NN will tend to repeat the correct answer should the test point in question be re-presented (see backpropagation in Chapter 7). The NN is adjusted during each iteration to improve a correct match (answer). The dotted learning curve in the figure describes the prediction learning path (that changes with each iteration) until it reaches a "plateau of acceptability." Continued learning cycles may only improve the chances of correct prediction only slightly, and expenditure of additional computational time may not materially benefit any changes. Once the NN adjustments are complete, it may be used to provide a judgment on data it has not previously seen: those of a student applying for advanced study. A new candidate's SAT and grades are presented to the system, and the probability of success will be generated.

This algorithm is similarly used to train a human to complete a task. (This is characterized as instrumental or operant conditioning where voluntary behavior is modified by a series of rewards or admonishments [Poling, Carr, & LeBlanc, 2002].) A young person learning to play basketball is first placed near the basket and attempts to throw the ball into the basket. Once he or she can successfully do so, the trainee is moved further away from the basket and repeats the task to be learned. A 100% success rate is not required at each stage, only some "acceptable" number of successes. In Figure 12.13, only two parameters are considered for demonstration purposes. Most DL machines use a great many parameters, although too many may lead to overspecification errors. (The DL

designer [a data scientist] designates the type and number of parameters.) The selection of appropriate parameters and the number of samples to be considered are important considerations within the new expertise specializations known as data science.[10] (For example, the "suitability" of a parameter may be studied using Bayesian analysis, discussed above.)

Table 12.12 summarizes DL assessment (and exhibit patterns that humans may present as well) when learning a new task.

Table 12.12 Assessment of learning algorithms.

Case	Designation	Assessment	DL Example
1	Peak or plateau	Learning continues to improve during each iteration, but increases grow smaller or may deteriorate at a peak so additional iterations are not warranted.	Learning to identify a specific type of apple from among many types. There are many other examples: face, speech, sound identification, etc.
2	Hilly	Mountainous experiences can produce a "false" learning point. A first peak causes learning to stop, but continuation of iterations can lead to additional increases in achievement. In some cases, the test set is deliberately changed ("toggled") to test for such instances. The revised data can then result in either Case 1 or Case 2.	A DL system may be trained to identify printed material (letters, numbers, special characters). Training might produce excellent results (peak success) for a letter such as "g"—resulting in a local peak. But even better results might be possible had other forms of "g" been included such as "$\mathcal{g}$" (a modified form of the letter).

ARTIFICIAL GENERAL INTELLIGENCE

Consider an IA that seeks "better" goals under unknown constraints as opposed to one that seeks a fixed goal under relatively well-defined conditions; that is the singular ambitious goal of artificial general intelligence (AGI) (Goertzel, 2016). One of the great challenges of AI is for a machine to function—maintain its purposes and goals—in an environment where unknown (or unforeseen) events can occur. Humans are considered adept at functioning in such environments. The "intellectual" abilities of such a machine might be indistinguishable from that of a human being.

As the human brain contains some 100 billion neurons (16 billion in the cerebral cortex) and 100 trillion synapses (neuronal connections), existing computer technology is no match. However recently announced computer technology employing "quantum"

[10]With respect to DL, data science includes math and statistics methods as well as computer science.

concepts ("quantum computing")[11] based on subatomic particles might "shortly" lead to a machine that competes with (or exceeds) the human brain's capabilities. The following was part of a recent *New York Times* article[12] concerning ongoing research at Google, and the performance times are notable:

> The device, in three minutes, had performed a highly technical and specialized computation that would have taken a regular computer 10,000 years to work out. The achievement, if real, could presage a revolution in how we think, compute, guard our data and interrogate the most subtle aspects of nature.

With such specifications, we may look forward to a machine whose basic component organization matches that of a human brain. Moreover, quantum machines capable of being physically introduced onto the body (bloodstream) provide the promise of such developments as gene editing for the elimination of many diseases and a more controversial application of improving human intelligence.

It remains to reproduce the cognitive capabilities of the brain, including consciousness. As of 2017, there were more than 40 organizations hoping to achieve the goals of AGI. There is no agreed-upon definition of such a cognitive agent, but several researchers have suggested that a system hoping to achieve this end must be able to demonstrate reasoning (strategize, make judgments under uncertainty), represent knowledge (including common sense), plan, learn (with guidance and independently), communicate (using natural language), and use these skills toward common goals. Ben Goertzel suggested the following in 2007:[13]

> The ability to detect patterns in the world and in the agent itself, measurable in terms of emergent behavior of achieving complex goals in complex environments.

To these capabilities, cognitive scientists add imagination and autonomy. Many of these capabilities have been demonstrated individually as described in this text but not yet in total combination and certainly not at human levels. Some have suggested a critical test or series of tests as evidence of achievement, as noted in the sampling shown in Table 12.13.

---

[11]Quantum computers are based on the properties of atomic and subatomic particles. Such bodies can store more than one piece of information in a single cell. This vastly increases both the amount of storage and the time required to access this information. This results in a quantum increase in computer capacity.

[12]https://www.nytimes.com/2019/10/21/science/quantum-computer-physics-qubits.html?smid=nytcore-ios-share

[13]https://www.youtube.com/watch?v=A-dycsiRwB4

Table 12.13 AGI demonstrations.

Test	Criteria
Turing	Human and machine in conversation with a judge unable to distinguish between human and machine.
Coffee (Wozniak)	The machine enters an American home and figures out how to make coffee without assistance.
Robotic College Student (Goertzel)	An intelligent agent enrolls in a college-level course and passes a program that results in award of a degree.
Employment (Nilsson)	An agent demonstrates capability in an economically important job as well as a human.
Furniture test (Tony Severyns)	The candidate is required to unpack and assemble an item of flat-packet furniture (e.g., read instructions and correctly assemble the item, including all fixtures).
Mirror Test (Tanvir Zawad)	Be able to distinguish between a real object and its mirror reflection.

The items mentioned in Table 12.13 subsume human intelligence commonly measured by the IQ test. Does a computer have an IQ? If one does not turn on the computer and provide it with a program, then the answer is "no"; the IQ of a computer is "0." When however, the computer is supplied with a program to take an IQ test akin those used by humans, then the answer is "yes," as disclosed in at least one report. Researchers at the University of Science and Technology of China have studied this problem and demonstrated that a computer can outperform humans in the verbal reasoning portion of an IQ test (using DL).[14]

AGI Problems

Computer scientists specify classes of problems that IAs are expected to solve—generalizations of the specific examples provided in Table 12.13. These are referred to as "AI-complete" or "AI-hard." Solving these is equivalent to the general aptitudes exhibited by humans and exceeds the problems previously discussed in this chapter, that is, problems that have intended goals or purposes. These AI-complete problems can be characterized as follows:

> Expected to include such features as computer vision, natural language understanding, able to deal with unexpected circumstances, and solve any real-world problem.

[14]htt ps://www.sciencealert.com/a-deep-learning-machine-just-beat-humans-in-an-iq-test

Currently, computer technology has not demonstrated such capability. In addition, it also requires human intervention and computation and is not "free-standing."

Computer-based systems have demonstrated many of the capabilities noted above but have not yet reached comparable capabilities equivalent to those of a human being: computational creativity (Veale & Cardosa, 2019), automated reasoning (Liò & Zuliani, 2019), decision making (Wang & Hammer, 2015), robots (see Chapter 13), evolutionary computation (Chapter 13), and intelligent agents (Chapter 13). Important contributions include those of Ben Goertzel (Goertzel, Orseau, & Snaider, 2015) and Doug Lenat.[15] Several examples of computer capabilities in the arts have been explored by Ellen Winner (2018), including music, visual arts, fiction, and the creation of art (by computer). Currently, humans (e.g., creativity) exceed AI capabilities in the arts.

To date, AGI has received limited attention from the research community. The chief complaint is that "intelligence" is so complex that it will not be completely replicated soon. There remains a core of scientists who are active and whose work is revealed at various AGI conferences.[16] In addition to Goertzel, the "futurist" Ray Kurzweil (2005) maintains that AGI will be achieved between 2015 and 2045.

REVERSE ENGINEERING THE BRAIN

Neuroscience has been a longstanding interest for scientists who have endeavored to reconstruct a model and understanding of the brain. These reconstructions include the neuron as a singular cellular unit; classification of neurons according to electrophysiological, biochemical, and molecular types; neural connectivity between brain regions; and neural mechanisms supporting function and behaviors.

During most of the 20th century and into the 21st century, AI has been greatly influenced by improved understanding of the neuron, the organization of the brain, and the development of the digital computer, including increasing capacities of its components as well as advanced imaging technology such as magnetic resonance imaging (MRI) and functional MRI (fMRI). The result of such combinations has been new entities having capabilities reflecting those of the brain coupled with behavioral phenomena studied by psychologists and embedded in a human form (Chapter 13).

During this era, biology (i.e., study of human organisms and vital functions, including its chemistry and the processes that occur in a living thing) has also contributed much to the cognitive sciences. Starting around the middle of the 20th century (1938), new "revolutionary" biological experiments began to be seriously entertained, namely, to replace the nucleus of an egg cell with the nucleus of another cell and to grow an embryo from such an egg. The first attempt in 1952—to clone a frog—failed. But a frog was eventually cloned in 1958 by John Gurdon (en.wikipedia.org/wiki/John_Gurdon). Successful cloning of a sheep occurred in 1954 (Fridovich-Kell & Diaz, 2020), and numerous animal clonings have followed; in 2009, a camel was cloned. This "cloning"

[15]https://en.wikipedia.org/wiki/Douglas_Lenat

[16]http://www.wikicfp.com/cfp/AGI Conferences

concept was embraced by researchers in AI who considered the possibility of "cloning" a human brain.[17] Could we "grow" a human brain? If we can map the neural network of the brain, can we reproduce it either organically or inorganically, and would we then be able to reproduce cognitive activities?

Key to many of advancements has been the development of MRI and fMRI. It had its origins in 1971 and has recently enabled scientists to map the brain with[18] increasing resolution capabilities. In 2017, scientists at Massachusetts General Hospital were able to attain an image having a 100-micrometer resolution—objects just one tenth of a millimeter in total size. It is among the most detailed (MRI) scans ever taken of the human brain anatomy.

We do not yet have a model of the human brain that can reproduce consciousness. However, several developments have led to consideration of new methodologies in which to exploit AI. The underlying motivation is prompted by viewing the brain as a collection of electrical signals. If these could be reproduced, in theory, one could upload someone's mind into a computer, allowing them to live forever in a digital form of consciousness. Three distinct approaches have emerged: simulation, emulation, and an "organic" implementation.

Methodologies

Simulation is an imitation of a real-world process over time in order to study alternatives or contexts. Simulation is currently a key methodology for study of the multiscale brain. It is often accomplished by using computer technologies whose architecture is built on an enhanced "Von Neumann" configuration consisting of a central processor, memory, and input/output facilities.

Software—a program—is central to this approach and consists of an "equivalent" model of the brain, which permits researchers to test capabilities of the system; the model mirrors human "thinking." Problem solving of this sort was explored above. This view emphasizes the importance of machine learning for constructing models. In most current machine learning applications, statistical and computational methods are used to construct models from large existing data sets, and those models are used to process new data. Examples include learning to classify spam emails, to enable fingerprint access to your phone, and to recognize human speech. An increasing number of machine learning applications do not stop learning after their initial training. They can either learn from additional data as they become available or even choose what additional data they would like to learn from. This last area is termed *active* machine learning,[19] and it promises to play an especially important role in biomedical research in the coming years. The process by which simulation has evolved is shown in Figure 12.14. Simulation and DL were noted above.

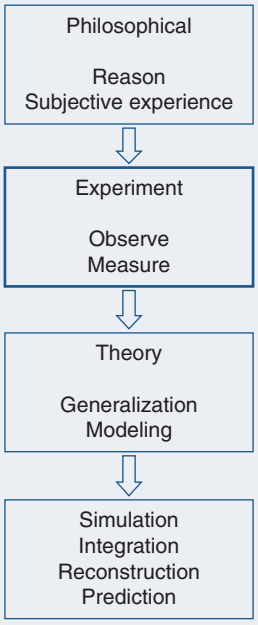
---

[17] https://www.wired.com/2013/08/ibm-modha-brain-language/

[18] https://www.sciencealert.com/100-hour-mri-marathon-gives-the-world-its-closest-ever-3d-view-of-the-human-brain

[19] Wikipedia Active Learning (machine learning).

Figure 12.14 Evolution of brain simulation.



In 2005, neuroscientists led by Henry Markram founded the Blue Brain Project, to which IBM contributed a Blue Gene supercomputer, at the Swiss Federal Institute of Technology in Lausanne. In 2013, the European Commission awarded Markram $1.3 billion to pursue an audacious goal: building a simulation of the human brain. The Ecole Polytechnique fédérale de Lausanne (EPFL) is a research institute and university in Lausanne, Switzerland, that specializes in natural sciences and engineering. It is currently the home of the Blue Brain Project, whose aim "is to build biologically detailed, digital reconstructions and simulations of the rodent brain and, ultimately the human brain."

Emulation, on the other hand, endeavors to equal or exceed the operation of a system (or a behavior) and consists of creating an exact model of the brain—neuron by neuron, including its interconnections: a "carbon copy." In this case, the "program" is inherent in the model, and the data presented to the model determine the results. This has led to the idea of a "brain on a chip." Piecing together molecules, neurons and brain regions will lead to understanding the brain. A dramatic event took place in 2014, providing encouragement for this objective.[20] In 2014, a collective called the OpenWorm project mapped all 1,000 cells, 302 neurons, and their interconnectivity of a worm

[20]https://www.sciencealert.com/scientists-put-worm-brain-in-lego-robot-openworm-connectome

(*Caenorhabditis elegans*[21]) and reproduced it on an integrated circuit. Moreover, of greater import, the scientists then attached the worm's simulated "brain" to a robot (Lego robot). The "worm robot" also included the worm's equivalent (limited) body parts: sonar sensor for the "nose" and motors on each side of the body as worm motor neurons. Without any additional software programming instructions, the robot responded to stimulations of the "nose" sensors. Stimulation of the nose (in the front) stopped the robot's (worm's) forward motions. Contact with the anterior and posterior sensors made the worm move either forward or backward. A sensor connected to detection of food made the worm move forward.[22] No neural training was required to achieve these results.

The mission of the European Union's Human Brain Project is to achieve an electronic model of the brain with functions reproduced in silicon (i.e., chips). One such neuromorphic[23] formulation is shown in Figure 12.15.

The very large-scale integration (VLSI) components in the diagram are integrated circuits that carry out the operations with their associated brain elements. The circuits emulate the neuronal structures like the worm model described above but with an extended number of neuronal elements. A system to accomplish this goal would need to carry out about $4 \times (10^{29})$ flops[24] and reproduce the actions of 90 billion neurons, 1,000 trillion synapses, 90 billion glial cells, and some 1 trillion molecules per neuronal cell.

The most scholarly, comprehensive study of such emulation is found in *Whole Brain Emulation: A Roadmap* by Sandberg and Bostrom (2008).[25] Some emulation developments are summarized in Table 12.14.

The "Organic Brain" and Wetware

In 1999, a remarkable event occurred when U.S. scientists from Georgia Tech and Emory University created a computer made of neurons taken from leeches was unveiled. (The neurons are harnessed in a petri dish by inserting microelectrodes into them. Each neuron has its own electrical activity and responds in its own way to an electrical stimulus and was able to perform arithmetic calculations as noted in a BBC report.[26]) In 2018, Andrew Adamatzky published a paper entitled "Towards Fungal Computer."[27] The

[21]*Caenorhabditis elegans:* a free-living, transparent nematode, about 1 millimeter in length, that lives in temperate soil environments.

[22]A video of this worm robot can be found at https://www.youtube.com/watch?v=PgSlIZCwyZs.

[23]Neuromorphic technology was developed in the late 1980s by Carver Mead using very large-scale integration analog circuits to mimic neurobiological structures found in the nervous system.

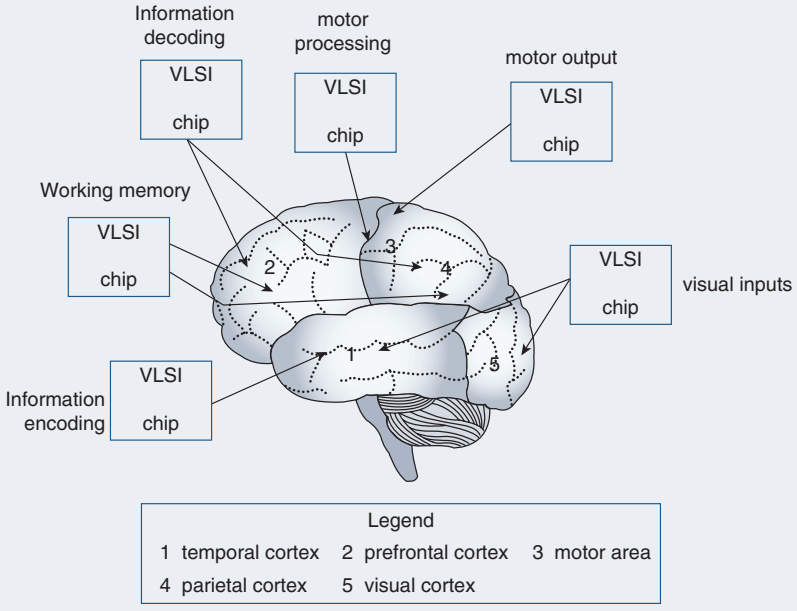
[24]A unit of measure for calculating the speed of a computer. The time required to complete one mathematical operation in a second

[25]*Whole Brain Emulation: A Roadmap*, Technical Report #2008-3, Anders Sandberg, Nick Bostrom, Future of Humanity Institute, Oxford University.

[26]BBC News, Sci/Tech, Biological computer born, Wednesday June 2, 1999.

[27]National Library of Medicine, National Center for Biotechnology Information, "Towards fungal computer," Andrew Adamatzky, *Interface Focus*, 2018 Dec. 6;8(6).

Figure 12.15 Wikipedia/Appaloosa from http://en.wikipedia.org/wiki/VLSI_Technology.



(a) **Sample "Brain on a chip" organization after Blue Brain Project**



(b) **Image of VLSI chip**

Table 12.14 Recent developments in emulation technology.

Year	Development
2006	GeorgiaTech; first case of a silicon array of neurons
2011	MIT: a computer chip that mimics the analog communication in a synapse between neurons
2013	Stanford University: NeuroCores; emulates 65,536 neurons
2013[a]	Human Brain Project: start of project to create a model of the brain using biological data (funded by European Commission)
2013	IBM TrueNorth chip/Obama Program: Brain Research through Advancing Innovative Neurotechnologies (BRAIN); dynamic understanding of brain function
2017	Intel: "Loihi" chip; used to implement learning and inference
2019 (pending)	IMEC (Belgium research center): first self-learning VLSI chip

[a]Goals: replication of brain architecture and activity on computers, medical informatics, brain-inspired computing, robots to test brain simulations, ethical and social implications.

abstract notes in part, "We propose that fungi Basidiomycetes[28] can be used as computing devices." This has come to be identified as "wetware," which seeks to address what are perceived as limitations in conventional computation: limitations of Moore's law and the restriction of two states in digital circuits. The wetware concept (as represented by the neuron computer above) is separate from the hardware of conventional computers and software, which is the instructions in memory units. Wetware makes use of organic modules such as the complex of cellular structures found in neurons. The molecules and chemical composition of the biological structure represent both the "hardware" and "software" of the entity. The "state" of a neuron is not limited to "1" and "0." It can shift between thousands of states by altering its chemistry and redirecting electrical pulses through more than hundreds of thousands of channels through synaptic connections.

Dennis Bray (2009) has explored this concept in the book *Wetware: A Computer in Every Living Cell*. Cells are complex computational structures similar to those of a computer. A cell can be viewed as a type of computer. Cells operate in tandem and receive input, process information, and compute an output. Each is akin to a computer on its own. The study of organic computers and wetware is an emerging subfield related to AI and is largely hypothetical. The creation of an organic computer developed by William Ditto and his colleagues at Georgia Tech signals the emergence of this technology.[29] A number of developments are summarized in Table 12.15.

[28]Fungus whose spores develop in basidia. Basidiomycetes include the majority of familiar mushrooms and toadstools.

[29]William, Ditto, PhD, labs.sciences.ncsu.edu

Table 12.15 Recent organic brain developments.

Event	Description
2014	Chemical computing system: international team from Swiss Federal Laboratories for Materials Science and Technology. Used surface tension calculations (Marangoni effect) to find the most efficient route between two points. It outperformed a satellite navigation system calculating the same route.
2015	Stanford University graduate students: create a computer using magnetic fields and water droplets infused with magnetic nanoparticles; illustrates basic principles of a chemical computer
2015	University of Washington students: a programming language for chemical reactions (DNA analysis)
2019	University of Glasgow; created a chemical computer using three-dimensional printing and magnetic stirrers to control oscillations of BZ[a] medium; perform pattern recognition
Ongoing	Eve Marder, Brandeis University: neural circuits in crustacean stomatogastic nervous system (30 neurons) can be reconfigured, using neuromodulators, to produce many outputs and behaviors without modifying the neuronal circuit. Dynamic clamp (conventional computer program): read electrical pulses from neurons in real time and interpret them. The program manipulated electrical signals to the neuron to represent numbers and communicate with each other to return an arithmetic sum.

[a]BZ medium: Belousov–Zhabotinsky reaction, or BZ reaction: nonlinear chemical oscillator includes a bromine and an acid.

ASSESSMENT OF AI AND AGI

The emergence of AI and AGI is the result of synergistic research between psychologists, physiologists, and computer scientists. Their efforts have produced remarkable IAs, devices capable of nearly perfect imitation of their human counterparts. A summary is shown in Table 12.16.

Despite these impressive examples, many researchers hope to achieve a higher goal, nothing short of the AGI discussed above. Ray Kurzweil sees the ultimate unity of the human brain and the AGI: the "singularity" of mind and machine. Kurzweil's forward-looking forecasts have provoked considerable academic criticism, particularly in regard to philosophical anthropology. The criticisms are summarized in a conference paper on virtual reality in a paper entitled "Pro and Cons Singularity: Kurzweil's Theory and Its Critics" by Bogdan Popoveniuc (https://www.researchgate.net/publication/266653924_Pro_and_cons_singularity_Kurzweil%27s_theory_and_its_critics).

Table 12.16 Things intelligent agents can do.

AI Facility	Capability
Read	Automatically read and summarize text; report back essential information, including a chain of records (blockchain).
Write	News media increasingly use AI for "standard" formulaic articles ("who," "what," "where," "when," "how"). Marketers craft social media posts using AI. An AI novel achieved literary recognition.[a]
See	Identify objects, people.
Speak	Maps respond to inquiries; schedule appointments; audio books; complete phone tasks in every conversational language; respond to human conversation.
Smell	Detect illness (aldehyde smell), including cancer and many other diseases; identify fumes (gas, caustic); perfumes.
Touch	Sensors (and cameras) to recognize product ripeness. (Others in development such as identify warehouse items)
Move	Vehicles, drones, robots
Understand emotions	Track human emotions in response to visual stimuli (videos); facial expressions; body language. (These are precursors of actions to be taken.)
Play games	Able to learn games without assistance and best humans at the activity.
Debate	IBM Debater can debate successfully against humans on complex subjects; create points of view; craft rebuttals.
Create	Make visual art; write poetry; compose music; take photographs; Google's "child" was successful in imitating a like human.
Read your mind	Interpret brain signals, create speech. Facebook, Elon Musk embarked on exploiting this possibility.

[a]"The Day a Computer Writes a Novel."

Scientists concern themselves with the possibilities of various extreme dystopian terminus. Even before a concept of AI entered the language, Norbert Wiener introduced the idea of an automatic, self-regulating control that he defined as "cybernetics." And he warned us of unforeseen consequences of this new technology in his 1954 book *The Human Use of Human Beings*, where he envisioned a time when the machine could become a dictator. The noted science fiction author Isaac Asimov proposed three laws that would "protect" civilization from this dystopian terminus:

- A robot may not injure a human being or, through inaction, allow a human being to come into harm.
- A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.
- A robot must protect its own existence if such protection does not conflict with the First or Second Law.

But clever writers and filmmakers have found ways to create intellectual content using these laws in unusual and counterintuitive ways, leading to dystopian conclusions. (The movies *I, Robot* and *2001: A Space Odyssey*, including the HAL 9000 computer, described as "foolproof," are prominent examples.) It would appear that human imagination cannot be matched, let alone exceeded, by machine intelligence. Can humans provide a set of unbreachable constraints on AI and, more important, AGI systems to avoid unforeseen destructive outcomes? AI and AGI technologies have extended both the potential for a new age as well as enhanced the possibility of creating an organic brain, an exact replica of its human counterpart.

John Brockman (2019) published the results of a project started in 2016 sampling the prospects for AGI from some of the leading researchers in AI. A sampling is shown in Table 12.17.

In 2016, Nick Bostrom—professor, Faculty of Philosophy, University of Oxford—published a revised version of his seminal book, *Superintelligence: Paths, Dangers, Strategies*

Table 12.17 Sampling of critical assessments of AGI/AI/super intelligence.

Contributor	Observation
Judea Pearl (Bayesian analysis)	We communicate with other humans and learn from them, instruct them, and motivate them in our own native language. We cannot communicate with opaque agents (AlphaGo). If we make changes in the task or operating environment, they will need to be retrained. New approaches are needed.
Stuart Russell (professor of computer science, UC Berkeley; author of most popular text in the field)	"Present AI studies the ability to achieve objectives, not the design of those objectives. Define a formal problem (F), design AI to be F solvers, so that no matter how perfectly a system solves F, we are guaranteed to be happy with the solution. If we have an appropriate F, we will be able to create provably beneficial results".
George Dyson ("Turing's Cathedral" Project Orion)	"Until we understand intelligence, we don't need to worry about superhuman intelligent machines."

(Continued)

Table 12.17 (Continued)

Contributor	Observation
Daniel C. Dennett (Tufts University, Center for Cognitive Studies)	"AIs should be treated as tools and not as humanoid colleagues." "AI (Strong AI) is possible in principle but not desirable. The AI that's practically possible is not necessarily evil—unless it is mistaken for Strong!"
Rodney Brooks ("Flesh and Machines: How Robots Will Change Us")	"The distinction between us and robots is going to disappear." Overly rapid development of software engineering has led to systems that do not have reliable safeguards.
Frank Wilczek (Nobel Prize in Physics)	"AI is not the product of an alien invasion. It is an artifact of a particular human culture and reflects the values of that culture." Computing develops → calculate the fundamental laws better and better → leads to better computer and the cycle repeats (this is similar to some ideas put forward by Ray Kurzweil).
Jaan Tallinn (codeveloper of Skype, Kazaa)	"Superintelligent AI is an environmental risk. We need to constrain a potential super AI from rendering our environment uninhabitable for biological life-forms."
Steven Pinker (Harvard, Johnstone family professor, cognitive psychologist)	"The abstract realm of knowledge, reason, and purpose can be linked to the physical realm of animals and machines via the concepts of information, computation, and control. A society with channels of feedback that maximize human flourishing can adapt to changing circumstances and can domesticate technology to human purposes."
David Deutsch (quantum physicist, Oxford University)	"Human-level AI provides promise for a better world rather than an Apocalypse." AGI should be "given its head," free to explore what other contributors consider dangerous.

(Bostrom, 2016), in which he warns us of the challenge presented by the prospect of a horrific superintelligence:

> [It] is quite possibly the most important and most daunting challenge humanity has ever faced. And—whether we succeed or fail—it is probably the last challenge we will ever face.

With respect to the issues related to control of an errant intelligence, he proposes a framework by which we may hope to obtain control of such a horrendous outcome. This system is summarized in Table 12.18.

Table 12.18 A framework for controlling superintelligence after Bostrom.

Control	Mechanism
Boxing	Access to world only through preapproved channels. "Caging" with limited access to sensors and information
Incentives	The system is limited to an environment of appropriate incentives. Social integration into a world of similar superintelligences Cryptographic rewards
Stunting	Impose constraints on cognitive capabilities—its ability to affect internal processes.
Tripwires	Diagnostic tests (without knowledge) with shutdown of dangerous activity
Domesticity	Limit agent's ambitions
Normativity	Rule-based principles specifying values to be pursued, particularly those considered to have ethical substance
Augmentation	Start with systems already imbued with human/benevolent motivations.

One outstanding element of the conversation remains to be resolved, namely, "what constitutes consciousness" in animals and humans. Such undertaking is just now emerging in the philosophical/psychological world. Its resolution will undoubtedly greatly affect the nature of any superintelligence that emerges.

Lest we think that the specter of a superintelligence has escaped the notice of the body politic and its laws, it is to be noted that U.S. lawmakers introduced a bill in 2019 that seeks to regulate AI. Titled the "Algorithmic Accountability Act",[30] it would require "big firms" to audit their machine learning systems for bias and discrimination and act if such issues were identified. It would also require these firms to audit all processes involving sensitive data for privacy and security rules.

Developments in the natural sciences have led to great understanding of the brain and how it functions, yet we are far from a complete description. Ideas about its operation have been shaped by technology, and technology has inspired new approaches to cognitive science. Historically, cognitive aspects of the brain have been modeled variously as a telegraph or telephone exchange. Currently, the supercomputer has received much attention as key to its understanding. The model has followed the technology of the age. As new hardware and software technologies develop (e.g., quantum computing, fMRI, organic computation), how our understanding and "reproducibility" of the human brain will inexorably progress for the benefit or detriment of humanity remains to be determined.

[30] https://www.congress/bill/116th-congress/house-bill/2231

SUMMING UP: A REVIEW OF CHAPTER 12

1. Artificial Intelligence (AI) is a composite of computational technologies that are inspired by—but typically operate quite differently from—the ways people use their nervous systems to sense, learn, reason, and act. Technological emulations of such functions comprise intelligent agents.
2. Human intelligence includes capabilities of the mind such as "learning," "problem solving," and "decision making." These subsume three logical processes: deductive, inductive, and abductive. They may be emulated on a computer that includes a central processing unit, memory, and elements to receive and report information. The memory contains a sequence of steps or program of operations to be completed.
3. Deductive logic: Logical sequences of IF–THEN rules lead the IA to reach a logically certain conclusion. Inductive logic: A logical conclusion follows from several premises that are found to be true all or most of the time. Abductive logic: A series of observations (and facts) leads to a conclusion that identifies a simple, most likely explanation of the facts.
4. Machine/deep learning: Using models of the neuron and cortex as well as statistical operations, an IA can detect patterns based on data that are presented. Data not previously presented can be used to improve accuracy and "learning" results. Such techniques are central to "data science." Humans are involved in the procedure because they choose the test data that the computer uses. If the data are flawed or biased, results will not be satisfactory.
5. Artificial general intelligence (AGI): The singular ambitious goal of an IA is to find an optimum solution to a problem under relatively well-defined conditions. Such an IA could rival human intelligence.
6. Modern technology has enabled the implant of computational elements into parts of the brain that may not be functioning. As new hardware and software technologies develop (e.g., quantum computing, fMRI, organic computation), our understanding and "reproducibility" of the human brain will inexorably progress, but the benefit or detriment to humanity remains to be determined.

SUGGESTED READINGS

Dreyfus, H. (1992). *What computers still can't do.* Cambridge: MIT Press.

Dworkin, B. R. (1993). *Learning and physiological regulation.* Chicago, IL: University of Chicago Press.

Hawkins, J., & Blakeslee, S. (2004). *On intelligence.* New York, NY: Times Books.

Reeke, G. N., Jr., & Edelman, G. M. (1988). Real brains and artificial intelligence. *Daedalus*, *117*(1), 143–173.

Searle, J. R. (1984). *Minds, brains, and science*. Cambridge, MA: Harvard University Press.

Service, R. F. (2014). The brain chip. *Science*, *345*(6197), 614–616.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis.* Cambridge, MA: B. F. Skinner Foundation.

Tang-Schomer, M., White, J. D., Lee, W. T., Schmitt, L. I., Valentin, T. M., Graziano, D. J., . . . Kaplan, D. L. (2014). Bioengineered functional brain-like cortical tissue. *Proceedings of the National Academy of Science*, *111*(38), 13811–13816.

Wiener, N. (1950). *The human use of human beings.* Boston, MA: Houghton Mifflin.

CHAPTER THIRTEEN

THE ARTIFICIAL INTELLIGENCE APPROACH II

Embedded Intelligence and Robotics

The idea of a mechanical embodiment in combination with a computational element was envisioned in the work of Alan Turing (Chapter 12). It is referred to as a "universal Turing machine" and consists of two parts: the algorithmic portion and a control mechanism. It could simulate any other control mechanism in accordance with the Church–Turing theorem, which deals with mathematical descriptions of those algorithms that could be implemented in a universal Turing machine.[1] Chapter 12 included topics important to the computational portions of such machines (e.g., state machines). This chapter will consider topics where the embodiment is highlighted but not to the exclusion of the computational elements of such machines when necessary. For example, the section entitled "Evolutionary Embodiments: The Merger of Human Cognitive Behavior, Biology, and Intelligent Agents" includes such topics. In behavioral circumstances the machine may not have any moving parts but reflect actions on the part of a user in combination with the computational elements (see "Machine–Human Interactions").

In 1994, Marvin Minsky, one of the key figures in artificial intelligence (AI), made the following comment when discussing the machine embodiment of intelligence[2]: "Will robots inherit the earth? Yes, but they will be our children."

Since then, we have been rapidly nearing his expectations as well as those of Ray Kurzweil's "Singularity," where brain and machine will be integrated entities.[3] Evidence in support of Marvin

Learning Objectives

After reading this chapter, you will be able to:

1. List several types of real-world applications of robotics.
2. Define the basic "anatomy" of a robot.
3. Describe a simple Braitenberg robot vehicle.
4. Discuss evolutionary algorithms in AI.
5. Contrast reactive and deliberative architectures in robotics.
6. Describe swarming robotics.
7. Describe Breazeals's approach to social robotics.

[1] https://en.wikipedia.org/wiki/Church%E2%80%93Turing_thesis

[2] https://web.media.mit.edu/~minsky/papers/sciam.inherit.html

[3] *Futurism*, October 5, 2017.

Minsky's provocative quote regarding the future role of robots in the world is growing daily. A few achievements of note:

- Remote medical consultation, evaluation, and delivery of drug treatment were conducted via an advanced robot with stereophonic hearing and a hypersensitive camera. Remote robotic surgery is being used more often (Field, 1996).
- Public agencies (e.g., military, police, and fire) have used such robots to neutralize or destroy bombs or carry out other dangerous missions. Increasingly, robots are being used in instances that can be described as dangerous, dirty, or dull—the three Ds—and the robots are required to "make decisions" on their own. Robotic control of anesthesia is becoming a valuable operating resource. Robotic surgery is a component of medical operations. Robots are "teaching" humans how to retrain their brain to compensate for neural impairments.
- NASA is making increasing use of robots for planetary mapping, where delays in communication make remote control virtually impossible.
- Use of robots in the home is on the rise (e.g., robotic cleaning).

In the 16 years since his original commentary, Minsky summarized what he observed regarding robots as follows:

> Some models have artificial intelligence that lets them do some things on their own, and they will inevitably grow smarter and more agile. They will not only represent human users; they will augment them.

Other examples can be cited: Computer scientists are developing machines that can teach people simple skills such as household tasks and vocabulary; AI elicits empathy from machines, robots, and devices developed to soothe, support, and keep us company. These are precursors to the possibility that human beings and machines will merge and overcome illness and perhaps death as we know it, advances in which machines can listen, speak, see, reason, and learn.[4]

Of particular interest are two recent research achievements: A robot learned locomotion skills by imitating an animal,[5] and Columbia University engineers conceived of a robot that can form a mental image of itself—a characteristic of self-awareness that has been a key limitation between machine and human.[6]

---

[4]The *New York Times* website (http://projects.nytimes.com/smarter-than-you-think/) provides a number of examples where AI and robotics are combined (e.g., "Aiming to Learn as We Do, a Machine Teaches Itself").

[5]techxplore.com/news/2020-04-animal-locomotion-skills-robots.html

[6]https://magazine.engineering.columbia.edu/fall2016/pushing-limits-robotics

MECHANICAL BEGINNINGS

Minsky's remark makes note of the human mind as a remaining scientific mystery. Yet humans have attempted to reproduce the brain in machine form for over 3,000 years. In literature, we find a novel (*Frankenstein, or the Modern Prometheus* by Mary Shelley). In the opera *Tales of Hoffmann* (Jacques Offenbach), the protagonist falls in love with a mechanical doll. Mechanical imitations of animals have spanned thousands of years. They include impressive mimicking of animals and humans that could consume food, play an instrument, and even play chess. A clever fake of the 18th century—"Mechanical Turk"—included a human chess master hidden beneath the board who moved the pieces via magnets. In Mary Shelley's novel *Frankenstein*, a scientist creates an artificial, hideous, humanoid creature.

The origin of the word *robot* itself comes from a 1923 play *RUR* (*Rossum's Universal Robots*) by the Czech playwright Karel Capek. Its meaning stems from Czech words for "slave," "forced labor," and "work drudge" and is relevant for many of its uses to this day. Robotic developments outpace the rate at which one can report concepts related to such automata.[7]

The first electrical autonomous robot of the 20th century is the work of the British neurophysiologist, cybernetician, and roboticist William Grey Walter. His first implementations, made in 1948 and 1949, were in the form of tortoises so named because of their shape and slow movements. They could explore an environment and, using phototaxis,[8] were able to find their way to a recharging station when power was low. In one experiment, he placed a light on the nose of the robot and put it in front of a mirror. The robot began "twittering and jigging like a clumsy Narcissus." Walter suggested that this might be evidence of animal self-awareness. His efforts were the results of his studies of brainwaves, which showed that local connections in the brain could give rise to complex behaviors.[9]

A robot will normally include the functions shown in Table 13.1, but the specific implementation can take a variety of forms. It is not easy to categorize them because each has its own features; they may vary in shape, size, capability, and, more important, purpose. A brief list of such machines is summarized in Table 13.1.

Some newer robotic examples defy belief. The giant Japanese robot "Gundam" stands 60 feet high and weighs 24 tons. It is mainly a demonstration robot. Robotic development of machines like those in the table is continuously ongoing.

EMBODIED COGNITIVE SCIENCE

The Design of Intelligent Robots as "Biologically Inspired"

Biology is the study of living organisms and includes such elements of life as the structure of cells, the basic units of life, growth, and natural evolution. From this natural science, we

[7]https://www.scmp.com/news/china/science/article/3038783/chinas-latest-robots-move-so-smoothly-scientists-dubbed-them

[8]Phototaxis: the ability of an organism to move toward or away from light.

[9]Walter also discovered a negative spike in the brain—contingent negative variation (CNV)—that occurs half a second prior to awareness of pending movements. This questions the notion of consciousness or free will.

Table 13.1 A compilation of robot types.

Type	Purpose
Aerospace, drones	Bird flying surveillance, planet explorers
Consumer	AI-powered robot assistants, toys
Disaster response	Searching for survivors, inspect damages
Educational	Algorithmic test platforms
Entertainment	Evoke emotional responses (theme park robots)
Exoskeletons	Enabling physical rehabilitation, mobility, endurance, capacity for heavy loads
Humanoids	Asimo; the type of robot most thought of as a "robot"
Industrial	Usually including a manipulator arm to perform repetitive tasks; Unimate is the forerunner of factory robots; often operate in coordination with humans; robotic constructions of bridges and roadways
Medical	Surgical
Self-driving cars	Robots drive themselves around
Telepresence	Permits a user to access a location remotely (e.g., "Alisa")
Underwater	Submersibles (e.g., "Aquanaut," "Ocean One")
Research	Intended for researchers and may also belong to the robotic forms noted in this table

have learned that new species develop through inheritance of suitable traits as well as regulatory systems for their internal environments and transformation and consumption of energy. All this has direct consequences for the study of IAs (Floreano & Mattiussi, 2008). Biology provides direct inspiration for us to be able to automatically find solutions to optimization problems that are quite complex. From the study of living organisms, we can improve mechanical design of the products we use, discover innovative computer software, and design robotic systems. (For example, using ant-walking models—where coordination of individual legs is achieved through both neural connections and the environment—if one leg is lifted, the force on all legs changes instantaneously—we can design more mobile "legged" or "wheeled" robots.) Evolutionary computational models follow directly from natural evolution (i.e., diversity, survival, inheritance, and a selection process) (Koza, 1992).

The Importance of Biology

Psychologists and neurophysiologists have provided numerous examples of biological models that cover an increasing range of complexity. From animal studies, we apply findings from foraging behavior, path integration, visual navigation (ants and honeybees), predator avoidance behavior (fish), maze traversal (mice), and the social

hierarchies of chimpanzees. Human studies also provide important insights for development of IAs. Infants seem to be able to make distinctions (such as differences between cats and dogs) that are often difficult to reliably replicate in an IA. Humans also provide examples for high-level intelligence: They can recognize a face under varying conditions of light, distance, and angle; physicians heal us by understanding and performing appropriate medical diagnoses; and even game playing gives us opportunities to develop logical strategies for the solution of many problems. Examples are virtually boundless, with daily new emphasis on studies concerning consciousness and language that some would maintain make us unique among creatures of the world.

ROBOTIC EMBODIED INTELLIGENCE

Embodied cognitive science subsumes studies that explain the mechanisms underlying intelligent behavior and comprise three main methodologies: (1) the modeling of psychological and biological systems in a holistic manner that considers the mind and body as a single entity, (2) the formation of a common set of general principles of intelligent behavior, and (3) the experimental use of robotic agents in controlled environments. Embodied cognitive science includes subjects related to the study and development of robots.

Alan Turing, the renowned English mathematician, computer scientist, logician, cryptanalyst, philosopher, and theoretical biologist, suggested that machines might need a humanlike body to think and speak, noting that

> it can also be maintained that it is best to provide the machine with the best sense organs that money can buy, and then teach it to understand and speak English. That process could follow the normal teaching of a child. Things would be pointed out and named, etc. Again, I do not know what the right answer is, but I think both approaches should be tried. (Turing, 1950)

(We should ignore his ethnocentricity.)

Defining and Describing a Robot

The International Federation of Robotics (IFR) was established in 1987 to promote, strengthen, and protect worldwide robotics activities. It promotes research, development, uses, and cooperation in all facets of robotics. It can be considered an ultimate arbiter of the definition of a robot.[10] IFR reports that "there is no single agreed definition of a robot," although many definitions include the outcome of a task that is completed without human intervention because a robot should include some "intelligence" of its own. While some definitions require the task to be completed by a physical machine that moves and responds to its environment, other definitions use the term *robot* in connection with tasks completed by software, without physical (mechanical) embodiment. Robots may best be defined by their qualities; these are summarized in Table 13.2.

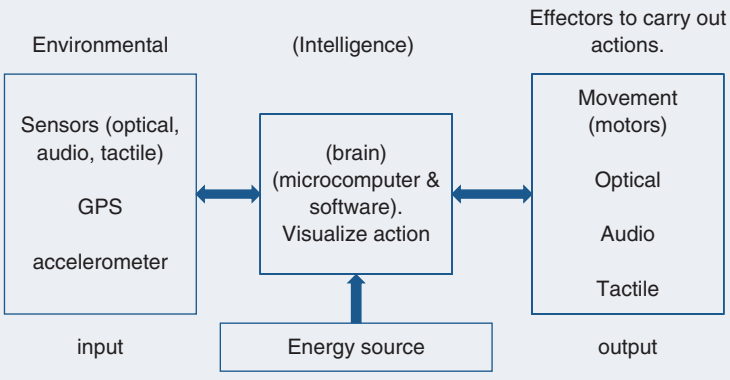
[10]The IFR supports the International Organization for Standardization (ISO) definition 8373 of a robot.

Table 13.2 Robot qualities.

Quality	Description
An artificial device that can sense its environment and purposefully act on it in that environment.	Describes robotics in terms of interactions with the world as shown by words *sense* and *act*. (It also encompasses devices that are under some human control.)
An embodied AI	The "AI" is an "intelligent agent" that provides the robot—possibly a machine—with its "cognitive abilities," without which it would be a useless shell.
A machine that can carry out useful work autonomously.	A robot should be useful; suited to efforts that are characterized as dirty, dull, and dangerous but not limited to such uses.

Robots are not in any sense "truly alive," although they appear to mimic humans quite closely in some circumstances. For example, they cannot autonomously sustain themselves; they need to have energy (although so do humans in the form of sustenance). Generally, they cannot physically grow or heal themselves. They cannot yet fully "build" copies of themselves (reproduce). If they achieved such accomplishments, we would need a new understanding of "being alive." Note that exceptions to these limitations can be cited; Ecobot[11] is a robot that can extract energy by "digesting" food. The study of robotic phenomena is as complex as the study of AI to which it is closely related. Total autonomy of this IA is a current goal of developers. A contemporary generic model of robotic architecture is shown in Figure 13.1.

Figure 13.1 Contemporary anatomy of a robot.



[11]https://en.wikipedia.org/wiki/EcoBot. EcoBot is short for "Ecological Robot" and refers to energetically autonomous robots that can remain self-sustainable, including collecting their energy from waste matter.

Robotic developments include the creation of humanoids, which are robots with a mechanical form, including elements of the human body, particularly those that are close to the kinematics[12] of humans. This relatively new field of "developmental" robotics may lead us to building humanlike intelligence. The challenge facing roboticists is to create humanoids that have feelings, can think, autonomously acquire new skills, and discover novel solutions to problems. The tools that roboticists have incorporated into humanoids and their relationship to biological capabilities are shown in Table 13.3.

Table 13.3 Robotic functions and biological equivalence.

Function	Type	Biology Relevance	Robot
Environment	Vision, hearing, orientation, smell, touch, sensing (range), location sensing	Eyes, ears, vestibular system, magnetics (birds), nose, animal heat (snakes), echolocation, no long-range sensing apparatus	Cameras, microphones, accelerometers, electronic compass, electronic nose, infrared sensor, artificial skin and whiskers, ultrasound/radar, laser scanner, GPS, gyroscope (reference direction, stabilizers, autopilots)
Communication	Lights, sound, gesture (face, body) radio	Coloring, voice, facial expression, body movements	Light bulbs, speakers, motorized skin, WiFi, or Bluetooth
Positional	Wheels, legs, flying	Muscles, legs, wings	Motors, wheels, wings
Control	Gross, fine	Muscles, arms, hands, fingers	Motors, arms, grippers
Energy	Electric power	Digestive processes	Motors, arms, grippers
Intelligence	Computer	Brain, mind (perception and cognition)	Algorithms
Structural	Variable	Bones, cartilage, exoskeleton	Metal, plastics

THE INTELLIGENT AGENT PARADIGM

Intelligence has been defined in many ways because it relates to many populations. We do not completely understand comparative intelligence among animal populations. How, then, can we extend this to agent (robot) intelligence? "Intelligence" is further elaborated throughout this text.

[12]Kinematics refer to motions not causally related to causes (e.g., the position of a person's ankle without referring to the neural or muscular forces that create the motion).

In its broadest sense, we consider intelligence of an entity (e.g., agent) from several perspectives. In the first instance, we include characteristics identified as "human" intelligence. It embraces the range of characteristics of intelligence: the ability to perceive or infer information, retain the knowledge that results from an action, and adapt it to the environment (see Chapter 12 for additional discussion). Table 13.4 summarizes intelligence as applied to robots (IAs).

To achieve the attributes cited in Table 13.4, a robot needs to include biological-like modeling, including reasoning; imitate behavior; and solve problems or achieve predetermined goals. They may use knowledge acquired in their interactions with the environment or learn to adapt to situational changes. (Such knowledge may be acquired to some extent independent of a robotic setting.) The integration of reason, behavior, and learning in a "mechanical" entity is the keystone of a robotic intelligent agent. This integration is summarized in Figure 13.1. As shown in the figure, there are three distinct elements: an "intelligence" component and two "mechanical" portions. Intelligence components most often include some sort of computer and/or electronic elements while the actuators (outputs) are present to reproduce any actions introduced by the computer software. The hardware may also include sensors to provide the data needed by the intelligence to determine the nature of the environment, which is an essential element of the agent's perceptive ability. Taken together, they encompass an intelligent agent or IA. (The output portion may include a visual element such as a display device.)

We will describe the integration of AI and machines as embodied in autonomous entities. Emerging research and developments regarding the distinctions between humans and machines will be explored. New research seeks to examine where the human seat of intelligence ends and the machine begins.

Properties of an Autonomous Entity

An autonomous entity is an IA situated in an environment, exhibiting rational behaviors, without human intervention. (Some minimal human intervention may be occasionally evident, but the situational embodiment circumstances must apply.) This translates into a set of requirements as defined in Table 13.5.

Table 13.4 IA intelligence characteristics.

Characteristic	Features	Restrictions
Instinct	Agent determines appropriate responses to stimuli.	Environment is known with limit on unexpected events.
Learned	Able to learn from its experience, humans, and other agents. Able to adapt and increase range of behaviors (particularly from uncertainty)	Safe and "smart" responses in presence of humans

Table 13.5 IA properties and salient features.

Agent Property	Description	Salient Feature
Rationality	Produces proper behavior at appropriate time in goal-oriented circumstances. Considerations: best outcome and performance assessment	Able to produce humanlike "thinking" behaviors
Sufficiency	Independent action; no external (human) control; seek goals or survival	Autonomous
Sustainability	Retains viability within the defined environment	Persistence
Communicability	Interact with users, including other agents and humans	Provide information or control of other agents
Cooperation	Collective or individual solution to problems in an environment with associated ability to communicate. May have capabilities for deception to facilitate its own advantages (reward)	Work with other agents to obtain common goals
Mobility	Able to migrate between other networks or environments	Able to negotiate hard or soft media (intelligence)
Adaptability	Adapt to the environment; able to create modifications of mapping between sensors and effectors that demonstrates learned, intelligent behavior and that satisfies given constraints	Can learn

Source: After Jones (2009) and Russell and Norvig (2009).

In most instances, a robot includes computational resources (hardware and software) without ongoing interference from a human. It is sited in the environment (that may include humans) to determine the best possible actions bounded by some constraints to perceive and move, if needed, to achieve a set of goals. (We consider this to be a working definition of the term *robot*.)

Environments of Intelligent Agents

A robot's ability to estimate its current state (how it is physically embedded in the environment) is an essential component of autonomy. It must have adequate computational resources at its disposal to take an action within bounds, have a perceptive ability, and move, if needed, to achieve a given goal (Jones, 2009). The toxicological elements of the environment include the following (Russell & Norvig, 2009).

Observability: Visibility of elements in an environment are important (e.g., pathways in a maze are not visible all the time).

Dynamics: Static circumstances or do the actions of the robot change the environment?

Experience: Are environmental descriptors available to the robot prior to its operation or does contemporaneous perceptual information satisfy the objectives of the system?

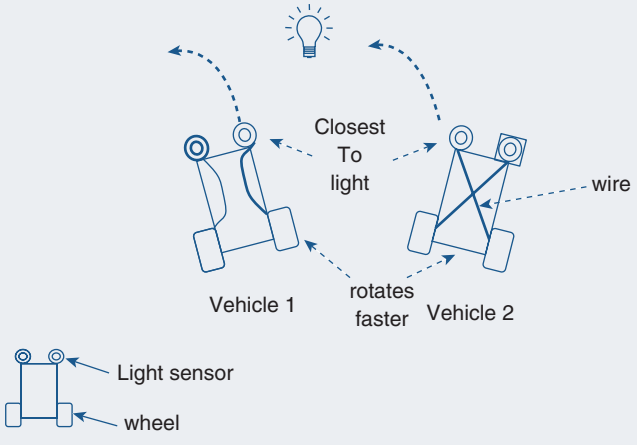
Continuity: Is the environment continuous—having a large (infinite) number of states—or discrete regarding the number of actions that the robot can pursue?

Other intelligent agents ("multiagent"): Does the environment require cooperative actions with other agents?

A Simple yet "Sophisticated" Robot

During his lifetime, Valentino Braitenberg served as the director of the Max Planck Institute for Biological Cybernetics. He was a gifted physician (as well as musician) and became interested in the study of the human brain in 1948. Despite recognizing their complexity, he concluded that computers were useful models for the brain. In addition to significant contributions regarding understanding the organization of the human cerebrum, he undertook research in cybernetics. In accordance with Wiener's ideas, he suggested that an analog vehicle—Braitenberg vehicle—using combinations of sensors and actuators could exhibit behaviors like human aggression, love, foresight, and optimism. These concepts involved connections between the structures and functions of the brain (Braitenberg, 1986). He suggested that depending on the interconnections between the sensors (e.g., heat, optical, tactile) and motor actuators, potentially different goal-oriented behaviors might result; consider Figure 13.2. In this formulation, each motor increases rotational speed with increasing signal intensity from the sensors.

Figure 13.2 Simple Braitenberg vehicles.



The right sensor in Vehicle 1 is closer to the light source and will produce a strong signal. As a consequence of the wiring, the right rear wheel will rotate more than the left wheel. Consequently, the vehicle motion will be away from the light, and the vehicle will seek a dark area. Applying analogous logic to Vehicle 2 will result in motion toward the light, a consequence of the alternate wiring. When additional light sources are added to the environment, Braitenberg vehicles will exhibit complex and dynamic behavior. Depending on the connections between sensors and actuators, a Braitenberg vehicle might move close to a source but not touch it, run away very fast, or describe circles or figures-of-eight around a point. While this behavior is strictly goal directed, it might appear to be intelligent, the way some minimal intelligence is attributed to say a cockroach. However, the functioning of the agent is purely mechanical with very limited "information" processing or other apparently cognitive processes. (There is no significant information "processing" in the definitional sense.)

The connections between sensors and actuators for the simplest vehicles can belong to the same side of the body (ipsilateral) or the opposite side of the body (contralateral), as well as excitatory or inhibitory, producing four combinations with different behaviors named fear, aggression, liking, and love. These correspond to biological positive and negative taxes present in many animal species.

EVOLUTIONARY EMBODIMENTS: THE MERGER OF HUMAN COGNITIVE BEHAVIOR, BIOLOGY, AND INTELLIGENT AGENTS

Serious challenges remain for the goal of robotic development built upon a biological foundation, with a full complement of human characteristics currently representing one of the most advanced approaches toward implementing AI. As noted in Chapters 7 and 12, deep learning (DL) depends on using a relatively simple neuron model in combination with a number of like elements to form a neural network. Neural nets have greatly expanded Norbert Wiener's cybernetic concepts, particularly with regard to achieving a desired target. The neural weighting factors are adjusted in prescribed ways (see backpropagation) to change the net's response. The achievements of these nets are measured against desirable goals and subsequently the weights readjusted to achieve improved scores presumably getting closer to the preset goal (e.g., face recognition, word recognition). Notice that the algorithm itself—backpropagation—is "fixed" (programmed). The result of the training algorithms is to establish a final input weighting of the neurons comprising the net. The limitation of this methodology is that it often does not totally reflect how humans think. Human learning is often characterized by "exploration," using curiosity, interest, satisfaction and dissatisfaction, pleasure and pain, boredom, excitement, or even "whim"—the so-called leap of intelligence. (Other global robotic behavioral models as described by Russell are described below.) Such cognitive responses were introduced in Chapter 1 and explored extensively in Chapter 8. In particular, discussion focused on why complex human behaviors emerged as a result of evolution of the human brain. Such evolutionary developments have carried over into robotic research and development. From evolution, we note that animal features might be enlisted for something not originally intended; feathers were not totally useful for insulation but later became handy for flight.

Challenges to DL and Its Algorithms

DL and its associated learning algorithms grew out of the "AI winter" of the 1970s when research was blunted by the failure of an early version of the Perceptron (Rosenblatt, 1962), which failed to live up to its expectations when it could not be trained to recognize many classes of patterns. The statistical gradient descent technique[13] employed by DL's associated learning algorithms has opened pathways to a great many successful applications. However, many challenges persist and are summarized in Table 13.6.

Table 13.6 Challenges to deep learning.

DL Challenge	Notes
Plausibility	Backpropagation does not exist in a biological neuron; DL can get stuck in a locally minimum solution.
Scope	If DL "sees" all possibilities, it will be perfect. However, it makes decisions on limit sets. When it generalizes (extrapolation), it is subject to error.[a]
Data hungry	Where data are limited, DL is often not an ideal solution. Humans readily extend inference that cannot be reproduced with DL.
Shallow	Limited capacity to transfer to abstract concepts[b] like "justice," "democracy," and "meddling."
Hierarchical structure	Sentences are treated as sequences of words. Sentences are built on smaller sentences. For self-driven cars, complex structure is needed.
Open-ended inference	Humans excel at combining sentences with background knowledge. DL systems do not accomplish this as readily.
Transparency	With billions of parameters in such systems, some may be difficult to identify; such variables as "last-character-typed" are identified within a network and not by themselves. Individual nodes not identified and "debugging"[c] is compromised.
Integrated	Not well integrated with prior knowledge. Training set is not "complete." Common sense not fully employed.
Confusion	Confuse causation and correlation. Such systems may analyze input–output relationships without consideration of causation.
Stability	Considerations of parameters in the world are often unstable (e.g., changing emotions of a humanoid or political circumstances).
Confidence	Systems may not be fully tested. Results are only "approximal" to human equivalent behaviors.

[13]Gradient descent: An equation that includes a number of variables is minimized using statistics to adjust the contributions of the variables.

DL Challenge	Notes
Boundedness	Results may only be applicable to short-term perspectives when intended for durable recommendations.

[a]The DL may conclude the following: If P is true → then Q is true. It also may generalize to the contrapositive: If (not Q) is true → then (not P is true), which can lead to error.

[b]In the game "Breakout," after 40 minutes of training, a DL machine decides that digging a tunnel through the wall is the best tactic! It has no idea of the idea of a wall or a tunnel, for that matter.

[c]Debugging is the process of finding and resolving errors or defects.

Emerging Tools

Several alternatives address the challenges noted in Table 13.6. These include the following:

- Coupling unsupervised learning in combination with DL (Stoean et al., 2020)—for example, using movies instead of labeled data. The changing nature of such information will provide a teaching circumstance as changes occur from frame to frame. Unsupervised environments provide opportunities to explore problem-solving situations such as observations associated with the behaviors of children building a Lego tower.
- Use GOFAI, which is good at symbol manipulation such as required for perceptual classification (e.g., identify objects). Combining GOFAI and DL is compatible with organization of the brain (which has been described in numerous chapters in the text) (Garcez, Lamb, & Gabbay, 2009). For example, the primary cortex area is akin to a DL system, while in Broca's area and the prefrontal cortex, we find higher levels of abstraction.
- Cognitive and developmental psychological research: Machines aim to replicate brain action and learn the mechanisms of the human even if they are not direct reproductions. What, for example, are the learning mechanisms supporting natural language understanding where humans currently have clear advantages over machines? Some areas of understanding include integration of "commonsense" knowledge, how it develops, and how it is represented (Chen, Tan, Kuntz, Bansal, & Alterovitz, 2020).
- Beyond such demonstrations as the Turing Test, require a system to watch videos, read, listen, and answer open-ended questions about contents. In addition, study game playing, and transfer learned strategies. Study infants to understand their cognitive processes. An embodiment (robot) that can build structures (e.g., assemble a piece of furniture from instructions) would

demonstrate significant progress in pursuit of both AGI and a humanoid that will be indistinguishable from its human creator. The goal reflects the integration of perception and inference.

Evolutionary Learning Algorithms and Intelligent Agents

Human intelligence emerged from biological evolution. Machine intelligence is humans' endeavor to re-create that intelligence. An IA with intelligence equivalent to that of a human is the mission of many AI researchers. What is unique about humans is when they are asked to solve a problem, they will apply relevant past experience but alternatively may find a solution based on a sudden desire or change of mind, especially one that is unusual or unexplained—in short, a "whim."

Evolutionary learning is not new. Alan Turing and John von Neumann noted that there are several algorithmic approaches to solving difficult problems[14]:

- Exploitative methods use direct (local) knowledge to produce results. GOFAI is representative of such methods. Problems that can be formulated as logical sequences will produce answers based on the truth or falsity of phrases within the algorithm and explored in Chapter 12.
- Alternatively, "exploratory" methods of analysis take stochastic approaches; humans celebrate their leaps into the unknown—their "whims." From computer science, we learn from the "free lunch theorem" that no algorithm exists that can outperform every other algorithm for every problem. Each problem requires its own domain-specific knowledge and solution. Stochastic algorithms are not restricted to specific domains and can explore previously unspecified alternatives to the problem solution.

The limitations and challenges previously enumerated are consequences of a limitation inherent in ANNs. The algorithms that have been discussed are based upon a "fixed configuration" of neurons and their interconnections. While the "weights" associated with the neurons are adjusted in accordance with the learning algorithm, the number and their interconnections remain fixed. The parameters (interconnecting weights) of the embedment are being optimized during the learning phase, but the number and interconnection of neurons remain fixed. In effect, the neuronal interconnections themselves define the algorithm. Can we alter the architecture—number and/or interconnectivity—of the net itself to improve overall outcomes? Specifically, we seek to find an improved algorithm during the learning process rather than optimizing the parameters of the fixed neuronal net. Evolutionary algorithm(s) (EA) might achieve this aim by mimicking natural selection. It provides new avenues by which AGI may be achievable just as heavier than air flight evolved once we stopped imitation birds and entirely new

[14]"Turing and von Neumann's Brains and Their Computers," brown.edu/Research/Istrail_Lab/papers/Istrail-Marcus012912FINAL.pdf, accessed September 4, 2020.

possibilities emerged. (Ironically, a recent bird robot demonstrated flight based on principles used for planes.[15]) Within AI, EAs become useful when the learning/solution phase of IA (robotic) problems requires time-consuming calculations (e.g., the traveling salesman problem[16]). Simply stated, EA is characterized by four steps corresponding to facets of Darwinian natural selection. In nature, a fit organism survives and proliferates, while unfit members die off and no longer contribute to the gene pool of subsequent generations. Table 13.7 summarizes the relationship between natural selection and an EA.

Figure 13.3 describes the process by which the EA proceeds and implements the steps annotated in Table 13.7. The generic methodology as described in Figure 13.3 includes entity evaluation, eliminating those entries below the truncation point, and mutation of the remaining population, thereby producing a new population. There are several variations of the basic algorithm, but detailed descriptions are not included in this text.

Evolutionary Computation[17]

Applying this concept to IA research is an emerging investigation area codified as "evolutionary computation." As currently envisioned, it is depicted by the following steps:

- Using a neural net, start by assigning random values to the weights between layers. (In general, this will produce very poor results as compared to DL or other algorithms.)
- Create a set of "random mutations"—"children" with slightly different architectures. (In its most advanced formulations, introduce mutations in the number and/or arrangements of neurons and connections.)

Table 13.7 Characteristics of an evolutionary algorithm.

Process	Description
Initialization	Create a population of solutions—arbitrary number of members, possibly randomly selected and/or centered about a suggested ideal.
Selection	Evaluate members using a fitness function (score).[a]
Crossover/mutation	Genetic operators; create new children from parental data; introduce new genetic material into the next generation.
Termination	Time limit or algorithm reaches a threshold of performance.

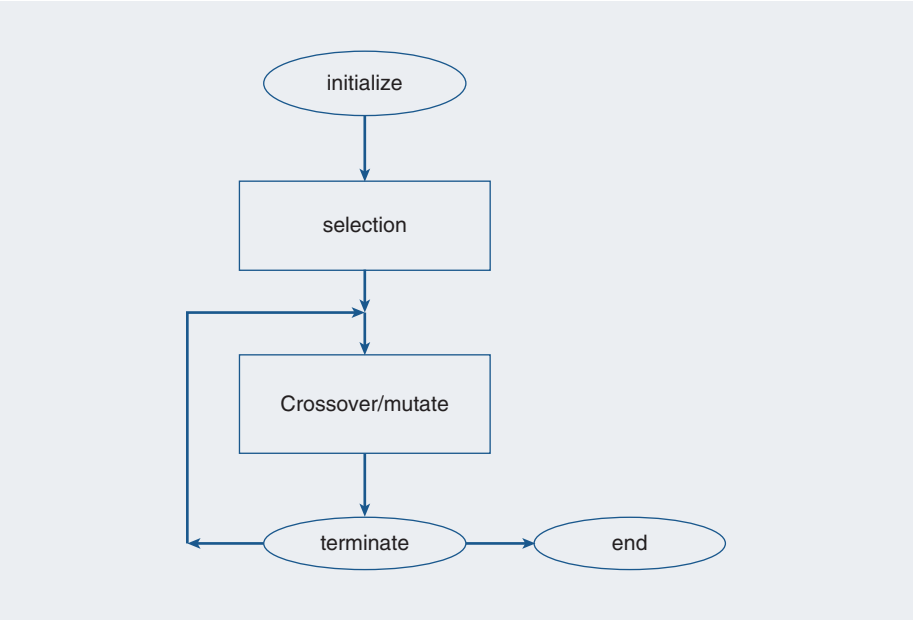
[a]Output number that evaluates calculates viability of member.

[15]"Drones Are Cool, but Robot Birds Are Better," Courtney Linder, *Popular Mechanics*, July 8, 2020 (accessed September 4, 2020).

[16]The traveling salesman problem: Given a list of cities and distances between each pair, what is the shortest possible route that (distance/time) visits each city and returns to the city of origin?

[17]*Evolutionary Computation*, 2011, 19(2), 189–223.

Figure 13.3 EA process.



- Evaluate the children's abilities.
- Keep the best; produce more offspring from the select children.
- Eventually, repeating the mutations will produce viable results. (The stopping point depends on cessation of generational improvements.)

The neuroevolution process is a "meta-algorithm," an "algorithm for designing algorithms" as opposed to the fixed (backpropagation) algorithm that defines DL. Such mutations are akin to what biological evolution has produced in humans. (It is to be noted that each generation may enter a "learning" phase to assess how much improvement is demonstrated during which backpropagation may be employed.)

The Evolutionary Mutation Process

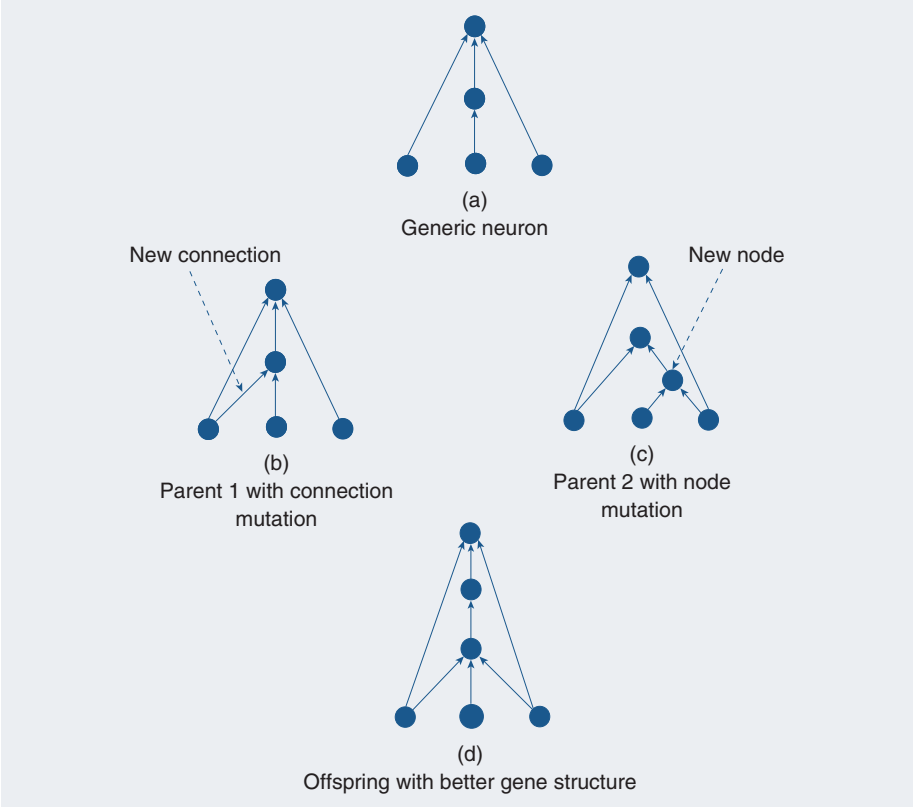
Figure 13.4 explains one step of the mutation process of an EA.

Figure 13.4A contains a simplified generic neuron that forms part of a DL network. In evolution, genes may mutate, and in the EA model, two possible (random) mutations are depicted. In Figure 13.4B, the mutation is accomplished by adding an additional connection to the structure while in Figure 13.4C, an additional node is added. These two structures are potential parents to the next generation (offspring) shown in Figure 13.4D. The offspring gains from both parents: both the additional connectivity of Parent 1 and the additional node of Parent 2 are retained in the offspring producing

an improved child. (This assumes that both original mutations represent improvements in the crossover/mutation process shown in Figure 13.3.) The offspring contains additional neural pathways and mimics a process that characterizes biological evolution, which produced the intelligence that marks *Homo sapiens*.

Machine learning employing an evolutionary algorithm is useful for obtaining good approximate solutions for problems where other methods fail. The EA is basically random in nature. It mirrors biological evolution, where changes in living organisms developed and diversified from earlier forms in Earth's history. The algorithm is characterized by its fundamental random nature and therefore never guaranteed to find an ideal solution but will lead to a good solution if one exists. For example, consider scheduling class timetables in a university. Such schedules must be compatible with room availability, faculty constraints, and interference with student schedules, to cite a few constraints. For humans as well as machines, it is "computationally hard." It is not efficient (i.e., time-consuming) to search for the optimal timetable because of the amount of computation needed. If some heuristics are used to define the problem, an evolutionary algorithm may produce a satisfactory solution. When NASA needs an antenna for a communication satellite, it

Figure 13.4 The mutation process (after Stanley & Miikkulainen, 2002).



uses such programs to find a design that satisfies several variables; the problem would challenge a human.

For such problems, a candidate solution is selected. Its performance is evaluated, and if it does not meet an acceptable criterion, the structure of the algorithm is changed. Random changes in the structure—the number of nodes or interconnections are altered randomly—are made and the evaluation is repeated until an acceptable solution is obtained. (Some programming languages have evolutionary frameworks that are available—"open-source" features [Malik, 2020].)

Robotic EA Examples[18]

One of the key elements of mobile robots is their ability to walk over various terrains. (Think of an infant learning to walk.) Over several iterations or design generations, the robot's frame must be adjusted to optimize body structure and applied muscular forces. Early generations may not be able to step over a rock or navigate in differing terrains. Design may be accomplished by computer simulations rather than by building a relatively expensive mechanical model for each iteration. More sophisticated robots that have evolved may incorporate the workings of the inner ear as a mechanism of balance. The example modeled after the galago, or Senegalese bush baby—a tiny primate with powerful legs that can leap meters into the air—has the robot imitate animal performance based on the biological model that the designer repeats. In this sense, it remains a sophisticated imitation based on a well-defined algorithm.

A popular test for evaluating robotic (EA) problem solving is to place the machine in a maze and compare various algorithms for finding the exit. Traditional algorithms measure how close the robot comes to the exit by assessing its distance to the target as the crow flies. In 40 trials, such robots found the exit only 3 times; the 37 failures ended in a dead end. Using the evolutionary paradigm, the robot succeeded finding the exit 39 times out of 40 trials! It accomplished this result by making use of data that traditional algorithms jettisoned. In particular, dead-end failures were added to the evolutionary algorithm, providing additional information. (It learned from its mistakes.) A brief list of accomplishments includes predicting future city landscapes, design of an optimized floor plan, communication of antenna design, identification of defects in electron microscopy images, and forecasting financial markets. Of particular note of success is the "The AlphaGo" algorithm that defeated a master in the game of GO.[19] The algorithm consists of two operations: a tree-search procedure (see Chapter 7) and a convolutional network similar to that described here that guided the tree-search procedure after a DL phase of the game and its possible states. What remains a challenge is human understanding of the thinking processes that the algorithm followed in its successful pursuit of the strategies to win the game. One can cite the detailed instructions, but the conceptual strategies

[18]https://www.theverge.com/tldr/2019/5/24/18638461/jumping-robot-salto-uc-berkeley-bush-baby-saltatorial

[19]GO, the game, was invented in China more than 2,500 years ago and is believed to be the oldest board game continuously played to the present day. The number of possible board positions is around 10^{17}.

Table 13.8 Emerging applications of EA.

Application Family	Examples
Natural science, math	• Facial composites by witnesses • Feature selection for machine learning • Traveling salesman problem and applications (multivariable analysis)
Earth sciences	• Climatology—heat flux • Climatology—global temperature change • Water resource systems • Groundwater monitoring networks
Social sciences	• Antiterrorism systems • Linguistic analysis; natural language processing (NLP), grammar reduction
Biological sciences/ bioinformatics	• Bioinformatics; multiple sequence alignment • RNA structure prediction • Motif discovery • Gene expressions profiling • Clinical decision support • Computational neuroscience: conductances in biophysically detailed neuron models • Protein folding
Other	• Multidimensional systems • Mutation testing • Rare event analysis • International equity strategies

remain elusive. The machine, in effect, cannot "tell" the designer how to win the game. EA applications are enhancing or replacing DL for embedded intelligence (robotic) applicability (see Table 13.8).

ROBOTIC EMBODIMENTS

The breadth of robotic embodiment is exemplified by the following event. An article in *Forbes* magazine (November 7, 2017) included the following:

> On October 25, (2017) Sophia, a delicate looking woman with doe-brown eyes and long fluttery eyelashes made international headlines. She'd just become a full citizen of Saudi Arabia—the first robot in the world to achieve such a status. (Sophia was awarded citizenship at the Future Investment Initiative in Riyadh).

Recognizing a robot as a "citizen" is a recognition that some AI specialists argue is a forerunner of the time when humans and IAs—in the form of machines—will be indistinguishable. (It could create new issues; how, for example, would this equivalence affect laws regarding "robots" or humans, for that matter?) Intelligence, as discussed here, cannot be restricted to the abstractions of AI. It warrants the development of sensorimotor capability. This extends intelligence beyond the limited areas of brain such as the neocortex to include some of the oldest parts of the brain that reflect the millions of years of interaction with the environment. The evolution of the machine/brain entity is unfolding exponentially.

At the other remarkable extreme, the National Academy of Sciences (Kriegman, Blackiston, Levin, & Bongard, 2020) included an article entitled "A Scalable Pipeline for Designing Reconfigurable Organisms." The significance of this research is noted:

> A method that designs completely biological machines from the ground up: computers automatically design new machines in simulation, and the best designs are then built by combining together different biological tissues. This suggests others may use this approach to design a variety of living machines to safely deliver drugs inside the human body, help with environmental remediation, or further broaden our understanding of the diverse forms and functions life may adopt. The development exploited the advantages of evolutionary algorithms.

The research and experimentation followed the sequence shown in Table 13.9.

As the computer had predicted, the lifeforms that were produced—the robots—were able to move on their own (eventually) in a coherent manner and began exploring their environment after a few weeks. Researchers claimed they could push pellets around with spontaneous, collective organization.

Table 13.9 Evolutionary organic robotic development.

Step	Description
1	Using a supercomputer, thousands of potential life forms were created using a variant of evolution. The computer was tasked with finding designs that would work best for particular tasks (e.g., what was best with producing movement in a given direction). The computer tried hundreds of variants of simulated cells that a lifeform used to carry out such activity. It produced theoretical designs for the lifeforms.
2	Using stem cells from the embryos of African frogs, a microsurgeon created a lifeform from the design. The stem cells were incubated and then cut apart and reassembled into the design created by the computer. (A real organic material was used to create a lifeform that had never been seen in nature.)

The researchers anticipate that more complex versions of these xenobots[20] such as those with a pouch on their body would carry an object—a drug delivered by swimming through the body. They can regenerate and are biodegradable upon death. Additionally, what is claimed is that they would be able to repair themselves. (When almost sliced in half, they can put themselves back together.)

As with other aspects of the AGI discussion, the researchers note that unintended consequences are possible. Their complexity could grow to a point where humans would be unable to predict how they will start to behave.

ROBOTIC REALIZATIONS

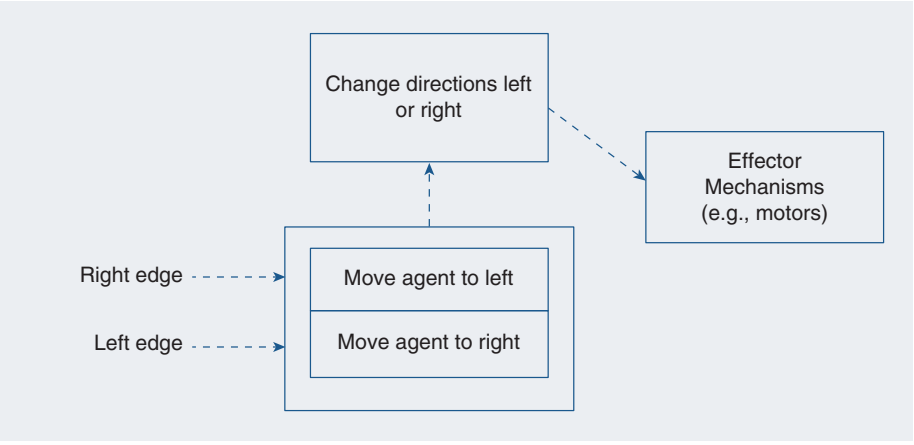
The span of embodiments noted, from the micro level to the humanoid level, embraces the models that govern how a robot might function in a real world. They subsume a number of frameworks akin to a building scaffold to support exploitation of machine embodiments. They include references, the capabilities, and resources related to sensing, perception, control, and effectors. While there are five primary architectures for robotic IAs, they fall into two broad arrangements depending on a key characteristic of its control elements (see Figure 13.1). These two arrangements are direct consequences of the development of computer architecture, namely, those employing a single processor as the essential element of the agent's intelligence and those employing a multiprocessor supporting the robot's intelligence.

- "Pipeline" (filtered) arrangement: The control element of the IA has a serial organization. Data are moved through a sequential system of processes that perform the transformation and/or mapping between the environment and the effectors. For example, a sequence might consist of perception followed by planning and terminating in an appropriate action. (This follows from the concepts originally introduced by Wiener.)
- Layered arrangement, which is characterized as a parallel organization. Various processes or actions are always available. The appropriate process or action is called upon as stimuli warrant. One action may supersede another if warranted. For example, if the agent is "exploring" and an "obstacle" is detected, exploration may be supplanted so that the agent can attend to the obstacle. Rodney Brooks is the originator of this architecture.

These two categories support five separately demarcated approaches to robot development (a fifth, hybrid combination architecture ["hybrid" architecture]).

[20]Synthetic lifeforms designed by computers to perform a function and built by combining biological tissues.

Figure 13.5 Simple line following or mapping reactive robot.



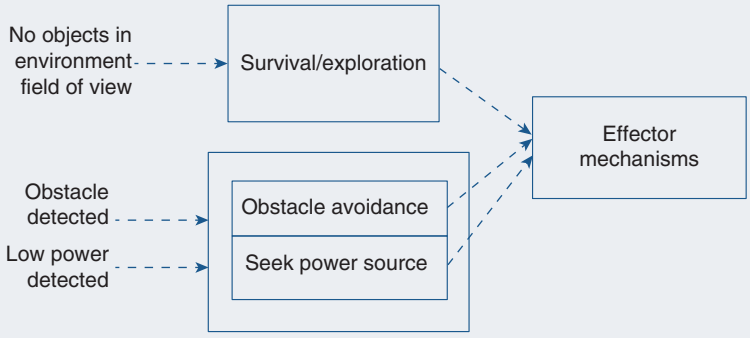
1. Simple reactive

A reactive architecture is shown in Figure 13.5. It is characterized by a direct mapping between sensor and actuator. This direct interaction between sensor and actuator produces an IA that is extremely fast and simple, as well as more economical with smaller IA space requirements. It has excellent rote learning properties as it has effectively "memorized" a list of responses. In reactive IAs, its response list and "reasoning" facilities are rooted in a series of IF–THEN logical sequences reflecting deductive thinking (Russell & Norvig, 2020). Its problem-solving capabilities rely on special-purpose solutions that pertain to well-defined environments. However, it has several limitations, including lack of extensive decision-making capabilities; agents are required to choose between competing stimuli with resultant unresolved confusion. It also relies on relatively simple environments as it may not adapt to the dynamics within its circumstances. This latter deficit limits its perceptive ability as it may not include sufficient internal processing to combine features and relationships of objects to detect unfamiliar objects. Figure 13.5 depicts a simple "line-following" agent. As the two stimuli (edges of a line) cannot occur simultaneously, the resultant actions will not interfere with each other.

2. Reactive/Subsumptive Architectures

Reactive/subsumptive EAs are more advanced forms of the reactive organization and have a "layered" arrangement. It is the result of the pioneering thinking of Rodney Brooks (1999). Advancing the idea of a "behavior-based" IA, he viewed the agent as a collection of simple behavior modules but having a hierarchical organization. (For example, "survival" behaviors are more important than "exploratory" actions.) An example of such architectures is shown in Figure 13.6. This agent indicates that there are two essential "survival"

Figure 13.6 Reactive/subsumptive agents.



Simple reactive/subsumptive organization for a freely moving autonomic robot. Control modules are concurrently available (parallel configuration) but only one module is in control of the effectors at any given time. A module may assume higher priority if necessary.

processes: If the agent detects a failing source of power, it must find replenishment or else it will "die"; if it has power but detects an obstacle with "life-threatening" potential in particular, it must take evasive actions. Lacking such threats, it may pursue other functions such as exploration (mapping) or foraging. They share many of the advantages of reactive architectures as well as some of its deficits. They can experience difficulties in changing environments such as circumstance not previously seen (e.g., staircase).

Brooks's initial demonstrations exhibited outstanding results. The robot, Allen—named in honor of Allen Newell, a key figure of AI—could do remarkable things at the time, primarily because of the speed at which it could travel (half a meter per second). It was built on a system of simple, primitive behaviors. It had no internal description of the world and hence could not "plan" complex actions. It could move down a corridor without colliding with walls, and it went around oncoming pedestrians. This became the platform for many other robots that included more and more capabilities (e.g., picking up objects). As a potential for a humanoid, it lacked emotions that is considered essential for such objectives.

3. Deliberative Architectures

As the name implies, this architecture uses the thinking processes akin to humans. Its fundamental elements include the current state and experience of the agent and the

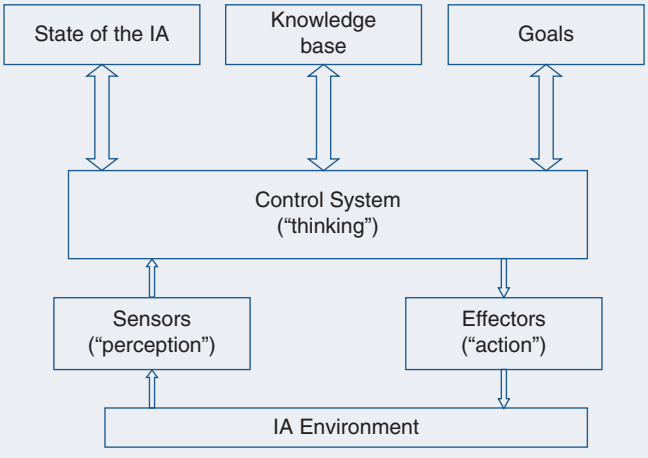
goals to be achieved. It has the potential to solve complex problems, includes provision for planning, and can perform a sequence of actions to achieve its goals. Because "thinking" requires processor resources, it can be considerably slower than other architectures. However, it has the potential to take advantage of contemporary AI technologies such as neural nets (NNs), fuzzy logic (FL), and expert systems (ES) to enhance planning actions. (Refer to Chapters 5, 7, and 12 for additional explanation of these methodologies.) Figure 13.7 depicts the deliberative architecture. One of the original ("classic") deliberative IAs was designated "Shaky."

The sequence of operations proceeds as follows:

- The environment provides information to the robot via the robot's sensors.
- Knowing its current state (e.g., position in the environment), the goals it seeks to achieve, and general rules that determines what it "knows" (knowledge base), the robot decides its next action.
- The robot's effectors (e.g., motors) respond to the commands provided by the robot.

The resultant action changes the environment, for example, the position of the robot relative to other objects in the environment. The entire process is then repeated until some terminal situation is encountered (the goal is achieved; the robot is stymied or encounters a catastrophe).

Figure 13.7 Deliberative architecture.



This model was developed at Stanford University, and its primary purpose was to create a robot that could make its own behavioral decisions (Nilsson, 1969). If it was given an instruction such as "move the block onto the table," the robot had to figure out how to do this (i.e., "deliberate"). For example, it had to

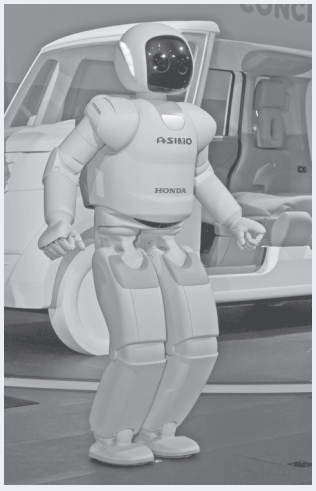
- Determine where the table was located and where the block was located
- Create a plan to move the block from its location to the table, including how to avoid any obstacles in the room
- To do so, it had to determine its environment, determine the next stage of the plan, and make the appropriate move.

Because of the relatively slow computer technology available in 1966, each of these steps took significant time: The result was that the robot proceeded in a discontinuous (or "Shaky") manner, hence giving rise to its name. Figure 13.8A shows a picture of Shaky while Figure 13.8B is an image of "Osimo," a greatly advanced "descendant" of the early "anthropoid." The ability to include many emergent AI concepts (fuzzy logic, NNs, DL) in such agents has made modern systems much better testbeds for IA research and development. (Impressive contemporary examples that "Osimo-like" robots demonstrate can be found at the MIT Biomimetic Robotics Laboratory.)

Deliberative architectures may take advantage of the many human intellectual capabilities. (These have been explored in greater depth in previous chapters.) They are briefly summarized here for convenience:

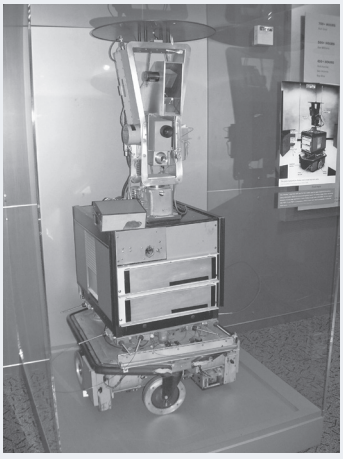
- Trial-and-error learning where responses to stimuli produce external errors that be rectified when such stimuli recur
- Rote learning: direct associations between stimulus and response that can be memorized from a list of such connections
- Operant conditioning: a highly developed form of learning involving positive and negative behavioral reinforcements
- Reasoning: ability to draw inferences: specific logical decisions may come from deduction, induction, and abductive thinking
- Problem solving: a variety of strategies for arriving at acceptable solutions to a problem, including, among others, special-purpose solutions, general-purpose solutions, and means-ends analysis
- Perception: creating a model (function) from observed facts and circumstances (or from internal stimuli such as dreams)
- Language: unambiguous knowledge representation and conversion into natural language representations

Figure 13.8 Osimo. (Consortium on Cognitive Science Instruction, supported by The Mind Project)



Source: Morio, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

(B) Embodied agents: Deliberative architecture examples.
(A) Shakey. http://www.computerhistory.org/collections/catalog/X279.83



Source: Carlo Nardone from Roma, Italy, CC BY-SA 2.0 <https://creativecommons.org/licenses/by-sa/2.0>, via Wikimedia Commons

4. Belief–Desire–Intention (BDI) Architecture

BDI systems store a representation of the state of the system, including the environment. This constitutes its "beliefs." In addition, it maintains agent goals ("desires"). And it contains a series of "mappings" for converting beliefs and desires into actions ("intentions"). The BDI architecture is within the broad realm of deliberative architectures (Wooldridge, 2000). Systems exploring closer relationships between humans and machines are discussed later in this chapter.

COOPERATING INTELLIGENT AGENTS AND SWARMING

Swarm behavior is the result of the interactions of group participants. Human activity often requires coordinated action. Examples include fighting fires or employees of a corporation working together to decide which products to develop or what to have for lunch. This is particularly true for team sports (volleyball, soccer). Sports can be studied by simulation of swarming behavior.

Animals exhibit swarm behaviors. Ants, for example, exhibit survival behaviors via traces left in the environment between agents (e.g., stigmergy). Such coordination can also be achieved via other types of communication. A trace (e.g., pheromone) left in the environment by an individual agent stimulates the actions or behaviors of a succeeding action either by the same agent or very often by different agents. The resultant swarming underscores survival, which is essential for successful evolution. An ant can find its way home or find food, and the same path is then available to other members of the colony.

Swarming Robotics

Nature provides us with many examples of advantageous structures and behaviors such as the swarming behaviors of ants, flies, bees, cockroaches, locusts, insect migration, bird migration, and marine life. (Erasmus Darwin observed that plant growth resembled swarming behavior.) Craig Reynolds (1987) developed a computer program to simulate such phenomena. It was based on a set of rules that characterized such behavior: Participants moved in the same direction, close to their neighbors, and avoided collisions with their neighbors. Such rules have delineated the taxonomy of the robotic emulations of these behaviors: special organization, navigation, and decision making (see Table 13.10).

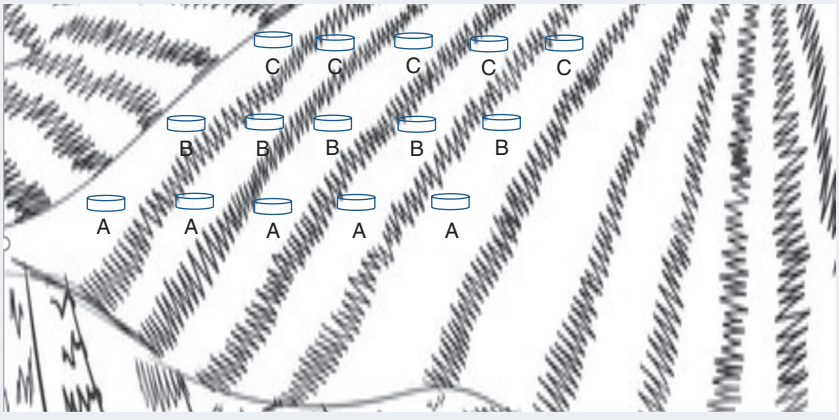
Smarming behavior modeling faces numerous challenges. Local interactions between elements are hard to predict. Essential communications between members are complex and confront industrial applications. In such cases, the term *swarm* may be used, but implementations are limited to centralized control, which is not an element of animal swarms. The taxonomy of swarm behavior is shown in Table 13.10.

Swarms have been studied for research and education. Specific examples include terrestrial, aerial, aquatic, and outer space. Swarming studies have been ongoing for several decades. However, those intended for industrial uses have lagged. One example of industrial swarming is shown in Figure 13.9, which depicts how swarms might be used in agriculture. The robots are configured to prepare the soil, disperse the seeds, or dispense fertilizer. Once the crop is ready to be harvested, a swarm can be employed to collect the yield. Other possibilities exist for drone swarms that can determine which plants require water or which are ready for harvesting.

Table 13.10 Taxonomy of swarm implementations.

Characteristic	Variables	Description
Spatial organization	Aggregation, pattern, self-assembly, clustering	Spacing for interaction, chain formations, self-assembly, manipulate objects (construction)
Navigation	Collective, coordinated motions	Collective exploration, coordinated motion, collective transport (move objects too heavy for one robot)
Decision making	Consensus, fault detection, perceptions, synchronization, group size, regulations	Swarm agreement regarding a single alternative Assignment of tasks to individual robots Combines locally sensed data into a larger picture Determination of deficiencies of individual robots
Miscellaneous	Healing, reproduction, human interactions	Swarm can recover from faults of individual robots Allows swarm to create new robots or repeat pattern created from many individuals

Figure 13.9 Schematic of swarm robot farming application.



Particle Robots: An Emerging Technology

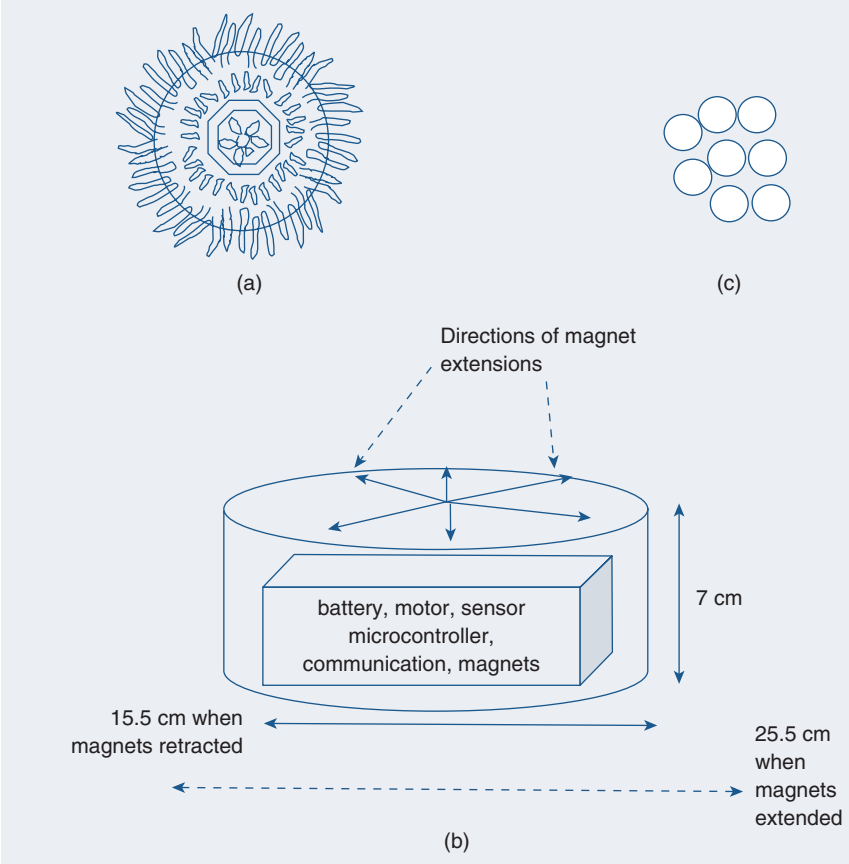
Biological cells include three important elements: the cell membrane, the nucleus, and, between the two, the cytoplasm. The cytoplasm, in turn, contains an intricate arrangement of fine fibers as well as organelles. When joined together, these cells comprise larger

organisms. For example, the sea urchin has a round body emanating spines (see Figure 13.10A). With this structure, the sea urchin can protect itself and, just as important, navigate in the water (https://arxiv.org/pdf/2003.08289.pdf).

Inspired by such natural phenomena, researchers at several institutions (e.g., MIT, Columbia, Harvard, and Cornell) have developed a unique robot that nominally appears to be incapable of any independent motion or locomotion and has no individual identity or an addressable position. This is at variance with robots that have been previously discussed that require explicit control of individual components (possibly from a central source) and perform specific functions; failure of a component threatens either the robot or the swarm in which it operates.

Each particle robot consists of a cylindrical base that houses a battery, a small motor, sensors that detect light intensity, a microcontroller, and a communication component that sends out and receives signals as well as magnetic attachments to interact with other particles. A schematic is shown in Figure 13.10B.

Figure 13.10 (A) Schematic of a sea urchin. (B) Schematic organization of the particle robot. (C) Particle group.



These loosely connected disc-shaped entities with minimum capabilities can push and pull each other and represent a new type of robot. Traditional robots usually have a single objective and include a great many parts. Failure of a single part can result in malfunction. With their simple design, a cohort of particle robots results in a more robust system. Failure of a single particle will not cause malfunction of the cohort, and adding another particle will not alter the operation. The magnetic extensions permit the units to connect and disconnect from each other, resulting in a variety of configurations and motion. In one demonstration of such a cohort, a moving light source results in a statistical mechanic coordinated motion through and around a variety of obstacles. Success was achieved even with failure of 20% of the particles. The resultant behavior would replicate the natural adaptability. They can explore spaces normally considered unreachable by humans. The coordinated motion can transport large object transport as well as movement toward a light stimulus. All of these are accomplished by cyclical expansion and contraction of the magnets in the particles. Many problems need to be addressed such as the ability to react to changes in tasks or the environment.

Embedded Intelligence as an Emotion Machine

Human emotions are a consequence of cognitive states that include subjective perceptions, conscious experience, a variety of psychophysical expressions, biological reactions, and cognition (mental state). By one counting, there are more than 75 different human emotions (e.g., affection, anger, angst, boredom, confidence, contempt, desire, depression, ecstasy, envy, self-confidence, sorrow, surprise, zeal, zest). Reproducing these behaviors in an IA is a challenge of extraordinary proportions. However, this has not discouraged—another emotion—numerous researchers from undertaking such endeavors. Two efforts in this regard are indicative of the emerging directions in robotic IAs.

Cynthia Breazeal can be considered a pioneer in creating robots imbued with emotions. She is a recognized leader in the development of social robotics and with Kismet, which first appeared in the late 1990s (Breazeal, 2002). "Jibo" is a more recent implementation of a social robot intended for use in the home. In a recent 2018 interview with the *IEEE Spectrum* journal,[21] she noted the following:

> "The way that the personal robots revolution is going to really happen is by making it a platform," she says. "Because once you do that, suddenly you can have a robot that can do many things for you, for many different people, versus these niche robots that only vacuum or only clean your gutter."

A summary of its resources and purposes of Jibo include the following:

- Interactive helper for families
- Equipped with an initial set of skills: camera capability (track faces, snap pictures, e.g., "selfies"), keeps track of family schedules, and helps people to interact

[21]http://ieeetv.ieee.org/ieeetv-specials/an-interview-with-cynthia-breazeal-ieee-vic-summit-2018?rf=events|125&

- "Storyteller" (relate stories with sound effects, graphics, movement)
- Will "play a role" in the social circumstances rather than being a "tool" as many such endeavors provide
- Open to developers: in keeping with current "app-oriented" social environments
- Recognize users and understand speech
- In addition to its internal processor and software operating system, it will include two cameras (detect and track people); microphone array (sound localization); body touch sensors, multiple motors, and belt system (for fast, smooth motion); and WiFi (for network connectivity).
- Normally plugged in for power but capable of battery operation for 30 minutes from an "app" installation such as those available for smartphones

Breazeal is confident that Jibo meets the definition of a "social robot." It may not seem like the extended idea of a robot as it does not have arms or legs, so its "human" image falls short of our prejudices.

MACHINE–HUMAN INTERACTIONS

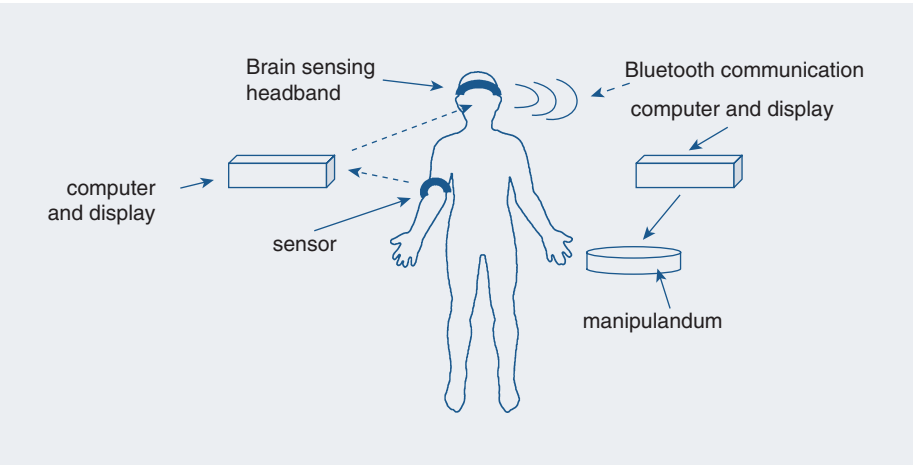
Ray Kurzweil has maintained that AI will lead to circumstances where the human brain and intelligent machines will eventually become indistinguishable. He named this concept "the Singularity." While considerable debate about the inevitability or date of this event can be found in the literature, symbiosis between man and machine has already been demonstrated in several applications, and new ones appear in the literature and media almost daily. In part, these are a consequence of the brain's plasticity. (See, for example, where an IA was asked to a write an essay for us from scratch. The assignment? To convince us robots come in peace.[22])

While the Singularity has not yet been realized, considerable understanding of the brain is now available to us. We know the brain interacts with and controls the functioning of every part of our body. In addition, it can affect living and nonliving elements using such phenomena as telepathy, telekinesis, and hypnosis. The schematic in Figure 13.11 conveys the ways machines have been able to interact both within and external to the body with the brain being a central element.

Two alternatives are shown in the figure. In one formulation, the waves that the brain emits are transmitted to a computer, where they are interpreted and translated into commands that a manipulandum can execute. For example, when the subject (human) wants to eat, the manipulandum can bring a utensil to the subject's mouth. In a mirrored possibility, sensors on the subject measure behavioral responses (e.g., energy from

[22]https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

Figure 13.11 Schematic of the brain, body, and machines.



the EMG being expended from a muscle) and the resultant information converted to a form (e.g., optical) that can assist the subject when making movements. (This is further explored below.)

Brain waves

The human brain emits complex electrical oscillations called brainwaves. The "power" of such signals from neuronal activities is small but reflects our thoughts, emotions, and behaviors. They differ in frequency range and express differing states of mind and body. These frequencies have been associated with deep sleep, drowsiness, calmness, alertness, cognitive functioning, and learning. (Sleep, as well as its relations to psychological states, has been studied for decades.) They can affect living and nonliving objects (telepathy, telekinesis, hypnotism, and even love). Historically, magnetic resonance imaging (MRI), functional MRI (fMRI), computed tomography (CT) scan, and X-ray studies have been used to collect brain activity. The Super Quantum Interference Device (SQUID) is among the most advanced of such systems. It is an extremely sensitive magnetometer that can detect extremely subtle magnetic fields. This makes it a powerful tool for biological studies. In general, such machines are extremely large, are expensive, and need technical expertise to operate. These machines have wide usage in medical diagnosis or specialized biological studies.

Advances in technology have brought us the ability to capture our own brainwaves using highly portable, wearable devices. They have been used to measure heart rate, body temperature, respiratory rate, and muscle tension, among other parameters. They have been used to deal with stress and improve focus. (The popular commercial

"self-help" feedback schemes have been the subject of much criticism.) A sampling of these wearable (headband) devices is listed in Table 13.11.

The Plasticity of the Human Brain

The *New York Times* featured the article "How the Brain Can Rewire Itself After Half of It Is Removed" (Sheikh, 2019). It described research on children who had required anatomical hemispherectomy—removal of half of the brain. Researchers reported and

> found that while the type of connections remained the same in the individuals with just one hemisphere, different regions responsible for processing sensorimotor information, vision, attention and social cues strengthened existing connections, communicating more frequently with each other compared with ordinary brains.

The brain was able to make use of previously underused neurons to restore behaviors never thought possible. It is but one example of the extraordinary plasticity of the human brain. As shown below, this plasticity can be exploited on a much broader scale, as noted below.

Machines Can Teach Humans

Human behavior modification has been the object of study in the psychological community since the early part of the 20th century (Edward Thorndike, law of effect). B. F. Skinner greatly advanced the underlying model by demonstrating the behavioral changes that were possible using "operant conditioning."[23] Such techniques make it possible to rehabilitate individuals who have suffered some insult to brain (e.g., lesion due to stroke). In such circumstances, the neural pathways and muscles may be functional, but taxonomy has been compromised by the brain lesion. For example, upper limb

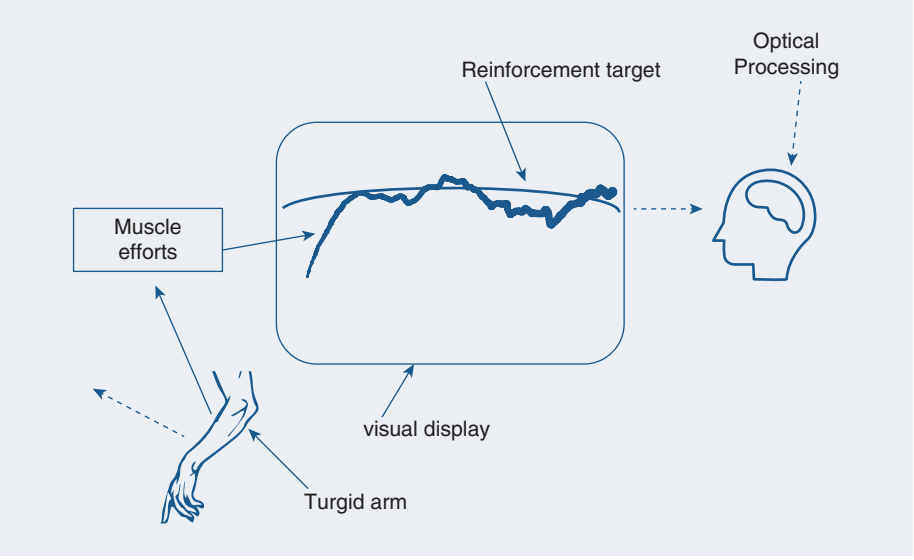
Table 13.11 Samples of wearable brainwave technology.

Headset Type	Notes
Thycn	Startup (2015); neurostimulation, targets autonomic nervous system, uses electromagnetic impulses to stimulate cranial nerves on face and back to improve mood
NeuroSky MindWave	Adapts electroencephalography (EEG) and electromyographic (EMG) technologies for the consumer market; works with a library of applications and programs for entertainment (toys and games), education, automotive, and health
Versus	Most advanced; brain training and biofeedback

[23] https://www.simplypsychology.org/operant-conditioning.html

paralysis occurs in people who have suffered stroke because both biceps and triceps are engaged ("fire") without coordination, and a characteristic paralysis follows. The biceps are physically "stronger" than the triceps, producing a turgid arm position. By making use of other parts of the brain's processing systems (e.g., visual, auditory, tactile), it may be possible to "retrain" these systems to replace the affected neuronal (motor) structure. This concept is shown in Figure 13.12. It employs "sensory substitution" in combination with "instrumental conditioning" (Silverman & Brudny, 1990). Instrumental conditioning is a combination of operant conditioning interfaced with robotic and/or intelligent agents. It is another example of the continued union of a cognitive agent and a human.[24]

Figure 13.12 Sensory substitution.



OVERALL EVALUATION OF EMBEDDED INTELLIGENCE

The world that we inhabit has grown increasingly complex, and to an increasing extent, we have come to rely on machines to make efficient use of our material and human resources. While some express fear that the economic goal of machine intelligence is to replace human workers with machine equivalents, we should remain aware of some attractive prospects afforded to us by such machines. A recent opinion in the *New York Times* described brain implants that could change humanity. It noted that brains are "talking to computers, and computers to brains."[25]

[24]https://www.jogohealth.com

[25]https://www.nytimes.com/2020/08/28/opinion/sunday/brain-machine-artificial-intelligence.html?referringSource=articleShare

In some activities that humans believe are intelligent (e.g., chess playing, computing math problems), machines may already outperform humans. However, the problems that weigh on replacing a human worker with a machine remain challenging—achieving the abilities of humans. Work on such problems proceeds at an accelerated pace. The *IEEE Spectrum* journal recently summarized the 100 most accessed articles over the past 10 years.[26] Additional applications have appeared in the news media with uses that affect our lives (e.g., computers can help speed up the diagnosis of brain tumors during surgery; robots can conduct an orchestra). A small sampling of the exciting practical opportunities within EAs includes the mining of information (i.e., data identification and interpretation), resource allocation, uncovering relationships that exist between scientific variables or rules that describe those variables, developing effective control and problem-solving strategies, improving the design of products that we use for activities of daily living, explaining decisions of all kinds, and identifying risks. Teaching humans is an exciting, emerging research and development area of inquiry. (Robots can teach young people programming!) These uses extend to the practice of medicine, commercial enterprise, law, social organization, and scientific investigation. The "intelligent computer" can be of enormous help to us. Such techniques as evolutionary learning and DL are making these possible.

Pursuing the study and development of machines with intelligence comparable to those of a human has another positive outcome: The more we try to replicate human intelligence, the more we may learn to appreciate and understand human intelligence. We need to consider several goals for such machines: a clearly defined task, a useful purpose, extensibility, and a well-defined organization.

However, designing computer-based machines that are intelligent is not the same as building computers that simulate intelligence. We should not be obsessed with mimicking human intelligence. At the same time, we should not reject methods that science tells us are involved in human intelligence. Algorithms for solving such game problems as chess are artificial and not truly reflective of the human intelligence that is brought to bear in such situations. We cannot, therefore, claim that we have replicated human intelligence, although it can be said that the IA has exhibited rational reasoning in that it used logic (and perception) and acted rationally in accordance with a goal-driven function. (To date, the machine has failed to achieve replication of human intuition.)

A fundamental question remains with respect to the IA subdiscipline and AI as a whole. It has been exquisitely posed by George Lakoff (2008):

"[Can] thought be characterized by symbol manipulation systems?"

At its ultimate core, IAs have yet to exceed the capabilities of a highly sophisticated symbol manipulator—a machine capable of decidedly refined functionalism. It remains to be seen if the "singularity" dreamed of by Kurzweil, Minsky, Goertzel, Bostrom, and others will be fulfilled.

To demonstrate the possibility of conversations about interactions between robots and humans, it should be pointed out that in 1999, an organization called the American Society for the Prevention of Cruelty to Robots (ASPCR) was founded.

[26]https://spectrum.ieee.org/automaton/robotics/robotics-hardware/automaton-most-popular-stories

SUMMING UP: A REVIEW OF CHAPTER 13

1. Embedded agents are precursors to the merger of human beings and machines that can overcome illness and prolong productive lives. They have the potential to listen, speak, see, reason, and learn at least at a primitive level.
2. Robotic predecessors originated some 3,000 years ago.
3. Embedded intelligence is intimately related to biology, which provides direct inspiration for us to be able to automatically find solutions to optimization problems that are quite complex. From the study of living organisms, we can improve mechanical design of the products we use, discover innovative computer software, and even design robotic systems.
4. Embodied cognitive science subsumes studies that explain the mechanisms underlying intelligent behavior and comprise three main methodologies: (1) the modeling of psychological and biological systems in a holistic manner that considers the mind and body as a single entity, (2) the formation of a common set of general principles of intelligent behavior, and (3) the experimental use of robotic agents in controlled environments.
5. The integration of reason, behavior, and learning in a "mechanical" entity is the keystone of a robotic intelligent agent.
6. The robot's ability to estimate its current state (how it is physically embedded in the environment) is an essential component of autonomy. It must have adequate computational resources at its disposal to take an action within bounds, have a perceptive ability, and move, if needed, to achieve a given goal.
7. Evolutionary embodiments: the merger of human cognitive behavior, biology, and intelligent agents.
8. Deep learning is the most popular algorithm for developing IAs. It has several challenges. The weights associated with the neurons are adjusted in accordance with the learning algorithm, and the number and their interconnections remain fixed. The parameters (interconnecting weights) of the embedment are being optimized during the learning phase, but the number and interconnection of neurons remain fixed. In effect, the neuronal interconnections themselves define the algorithm. We alter the architecture (number and/or interconnectivity) of the net itself to improve overall outcomes. Specifically, we find an improved algorithm during the learning process rather than optimizing the parameters of the fixed neuronal net. Evolutionary computation explores the possibility of modifying the neural net itself during the learning process.
9. Robotic implementations of EAs vary from the micro level to the humanoid level and embrace the models that govern how a robot might function in a real world.
10. Technology is approaching a point where an EA will be able to read and interpret the thoughts of a human.
11. Plasticity of the brain lends itself to reeducation by an EA using sensory substitution and the well-established psychological process of instrumental conditioning.

CHAPTER FOURTEEN

THE EMBODIED ECOLOGICAL APPROACH

A Dynamic Future for Cognitive Science?

EMBODIED AND EXTENDED COGNITION

For much of the 20th century, the cognitive and neural sciences generally assumed that the physical mechanism that underlies cognition is the electrochemical operations of neurons in the brain. Of course, different hormones and different sensory inputs can alter the function of that biological neural network. In fact, even microbes in the digestive system can alter the function of the brain. This "gut–brain axis" is involved in the development of autism and anxiety-depressive disorders (Carabotti, Scirocco, Maselli, & Severi, 2016). However, those alterations of the brain's function were not seen as constituting part of the mechanism of cognition itself; they were seen merely as external influences on that fundamental mechanism. Even functionalists (from Chapter 2) who tend to discount the value of neuroscience, favoring instead an abstract "software-level" description of cognition, generally assume that the brain is the sole physical hardware that runs that software in humans.

However, over the course of the first two decades of the 21st century, the sprawling discipline of cognitive science has begun to seriously consider the possibility that the physical mechanism that generates cognition may include more than just the neural networks of the brain. Rather than being located exclusively inside the central nervous system, cognition itself may be an emergent property of the seamless interaction between a brain and its body. Recall the discussion of robotic embodied intelligence in the previous chapter, where the design and operation of sensors and effectors play a fundamental role in generating robot cognition. It appears that your sensors (e.g., eyes, ears) and effectors (e.g., hands, mouth) may play a fundamental role in generating your cognition. The mental activity that comprises your mind may not reside solely inside your brain.

Learning Objectives

After reading this chapter, you will be able to:

1. Define *embodied cognition*.
2. Describe a laboratory experiment that supports embodied cognition.
3. Discuss the role of dynamical systems theory and "state space" in cognitive science.
4. Contrast classical cognitive science and dynamical cognitive science.
5. Define *ecological perception* and the role of affordances in cognition and action.
6. Define *extended cognition*.
7. Discuss the importance of interdisciplinarity in cognitive science.

Imagine grasping a crisp apple with the intent of biting into it. As your fingers close in on its shape, the pads of your fingers press into the surface of the apple and they depress somewhat because they are squishier than the surface of the apple. Thus, milliseconds before the nerve cells in your skin have begun to send electrochemical signals to your spinal cord to report on this feeling of pressure, the nonneural tissue of your finger pads is already conforming to the shape of the apple in a way that will assist in it being successfully grasped and lifted. In the field of soft robotics, where they design hand-like effectors with pliable rubber or foam, this squishiness is called passive compliance. Passive compliance is a form of **morphological computation**, where certain cognitive, perceptual, or motor processes are carried out not by the electronic circuitry inside the robot but instead by the morphology (e.g., shape and materials) of the robot's body. Via this passive compliance in your hand effector, your body itself is doing some of the intelligent sensorimotor negotiation with that apple, over and above what your brain is doing. For instance, if you were to develop thick calluses on your finger pads from hard work with heavy tools, that passive compliance would be substantially different, causing you to have to grasp that apple in a slightly different manner. Therefore, even with an identical nervous system, the sensorimotor intelligence inherent in your callused hand effector would be different, thus making your cognition a little different. This is just one example of how cognition may be seen as embodied.

PERCEPTUAL SYMBOL SYSTEMS AND MOTOR AFFORDANCES

Cognitive scientists have produced a wide variety of experimental demonstrations of how cognition is embodied. For example, the perceptual symbol systems theory, developed by Lawrence Barsalou (1999; see also Pezzulo et al., 2011), has motivated numerous behavioral, neuroscientific, and computational explorations of embodied cognition. In the perceptual symbol systems theory, it is argued that cognition does not use amodal logical symbols that are divorced from their analog sensory and motor impressions, but instead cognition uses those modal sensory and motor impressions themselves (e.g., from the visual modality or the auditory modality) as the objects of cognitive operations. By combining these perceptual and motor impressions in complex ways, the mind can generate **perceptual simulations** that are the currency of cognition—instead of amodal symbolic representations.

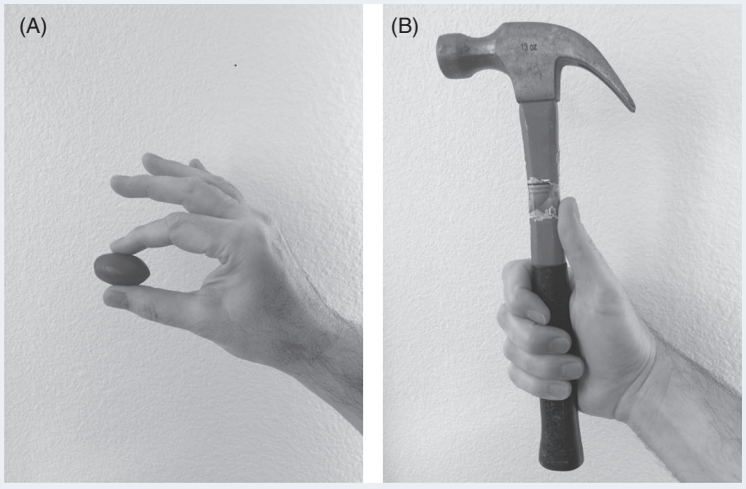
Recall from Chapter 5, where we described an experiment by Wu and Barsalou (2009) that asked a first group of participants to list the features of their concept of a watermelon and a second group of participants to list their features of half-a-watermelon. An amodal conceptual system would likely access the same conceptual representation in those two instances and produce lists that are basically identical. However, Wu and Barsalou's two groups of participants produced noticeably different lists. The second group listed the features "red" and "seeds" far more frequently than the first group. It appears that they were generating perceptual simulations of those

concepts with enough visual imagery to highlight the presence of the watermelon's seeds and red flesh for the concept of half-a-watermelon but not for the concept of watermelon.

There are now so many experimental findings in support of embodied cognition that entire books have been dedicated to the topic (Gibbs, 2005; Shapiro, 2019). For instance, Dijkstra, Kaschak, and Zwaan (2007) found that autobiographical memories of a childhood dentist visit showed faster recall when the person was sitting in a reclining chair than while standing. Conversely, autobiographical memories of childhood sports experiences showed faster recall while the person was standing than while reclining. Thus, memory was improved when the body was in a posture that matched the posture in the memory. Another influential experiment comes from studies of visual perception, showing that how one recognizes a visual object is not solely by processing its visual features but also by understanding what actions the object *affords* the observer (Gibson, 1979). For example, a grape-sized tomato *affords* a precision grasp with thumb and forefinger and also eating, whereas a hammer *affords* grasping with your whole hand and also swinging (see Figure 14.1). **Affordances** are best defined as potential action-based relationships between a particular organism and a particular object (or environment). To explore the role of these affordances in visual object recognition, Tucker and Ellis (2001, 2004) developed a manual device that allows a participant to respond to a visual stimulus either with their thumb and forefinger (mimicking a precision grasp) or with their palm and the three other fingers (mimicking a whole-hand grasp). They had participants view solitary objects on a computer screen (such as a grape tomato or a hammer) and categorize them as manufactured objects or natural objects. When participants were categorizing a grape tomato as natural, they were faster when they used a precision-grasp action than when they used a whole-hand grasp action. Conversely, when they were categorizing a hammer as manufactured, they were faster with the whole-hand action than with the precision-grasp action (see also Grèzes, Tucker, Armony, Ellis, & Passingham, 2003). These results suggest that even when making a relatively abstract semantic judgment about a visual object that is obviously out of reach, motoric properties of how one's body would interact with that object (i.e., its motor affordances) nonetheless play a role in the cognitive process. Hence, visual object recognition appears to be embodied.

Recognition of visual *events* is also embodied. Recall the discussion of the mirror neuron system in Chapter 11, where a region in premotor cortex becomes active while the observer is passively viewing someone else perform an action. Beatriz Calvo-Merino and colleagues (2006) recorded the neural activity of expert ballerinas in an fMRI scanner while they watched video clips of ballet dance moves. They found that part of how a ballerina visually recognizes dance moves that she herself has carried out before is by activating not only representations in visual cortex to perceive what she is looking at but also by activating representations in motor cortex to simulate how her own body would do the moves that she is looking at—even though her body is not currently moving at all. When the ballerinas watched equally familiar ballet moves that they themselves had not carried out before (because they are exclusively male ballet moves), only the visual cortex was active, not the motor cortex.

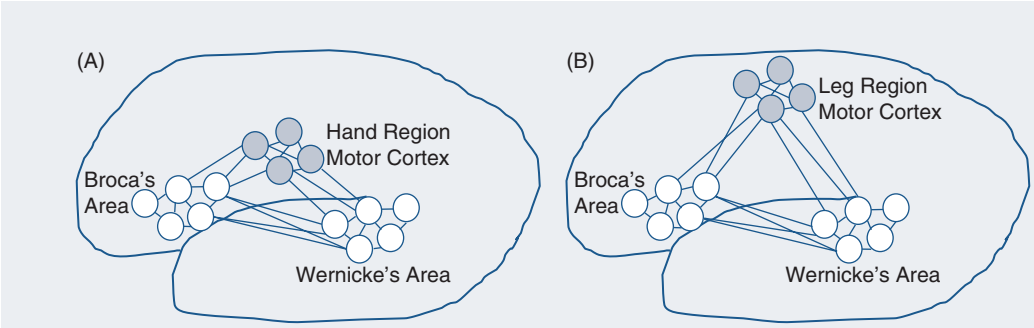
Figure 14.1 Precision grasp of a grape tomato (A) and whole-hand grasp of a hammer (B).



Now, let's consider language processing. If simply recognizing a visual object or event cannot help but take into account how your body would act on it, does recognizing a word also take into account your embodiment? Tatiana Nazir and colleagues (2008) had participants perform a simple reaching task over and over, but just as the reach began, an irrelevant written word was presented on the computer screen in front of them. When the word was an *action verb*, the acceleration profile of the reach revealed a milder deceleration about a third of the way through the reaching movement, compared to when the word was an *object noun*. Essentially, the visual input of that linguistic information (despite being irrelevant to the task) managed to subtly influence the temporal dynamics of the reaching movement but only when the information referred to human actions. Embodied cognition suggests that the neural representations of word meanings will be partially overlapping with the neural representations of the motoric and sensory processes that those word meanings refer to.

According to this account, not only should words influence motor processing, but motor activation should, in turn, influence word processing. Pulvermüller, Hauk, Nikulin, and Ilmoniemi (2005) demonstrated just that by gently activating motor cortex via transcranial magnetic stimulation (TMS) during a simple word reading task. If cognition was not embodied, then activating motor cortex should have no influence on

Figure 14.2 **Embodied neural representation for hand-action verbs (A) and embodied neural representation for leg-action verbs (B). When motor regions (in gray) are preactivated by TMS, word meanings get primed (Pulvermüller et al., 2005).**



language comprehension. However, if the neural representation of hand-action verbs (such as "catch") and leg-action verbs (such as "kick") includes not just neural ensembles in language cortical regions but also neural ensembles in motor cortex, then preactivating those specific motor cortical areas should improve the processing of those relevant word meanings (see Figure 14.2). As predicted, Pulvermüller et al. found that reaction times to hand-action verbs were faster when TMS had mildly preactivated the hand region of motor cortex. Conversely, reaction times to leg-action verbs were faster when TMS had mildly preactivated the leg region of motor cortex. Thus, the process of recognizing written words that refer to action includes, as one of its functional components, the activation of certain parts of motor cortex. Hence, language processing appears to be embodied.

Of course, externally imposed magnetic fields are not the only way to activate motor cortex. Simply carrying out a motor movement also involves activating motor cortex. Shebani and Pulvermüller (2013) had participants rhythmically move their hands or their feet while memorizing a list of words. If the hand region of motor cortex is partly involved in representing hand-action verbs, then trying to memorize those words while the hand region of motor cortex is instead devoting itself to a completely different task should disadvantage the processing of that hand-action verb. The same goes for the leg region of motor cortex and leg-action verbs. As predicted, Shebani and Pulvermüller found that rhythmic movement of the hands resulted in reduced accuracy for memory of hand-action verbs but not for leg-action verbs. Conversely, rhythmic movement of the legs resulted in reduced accuracy for memory of leg-action verbs but not for

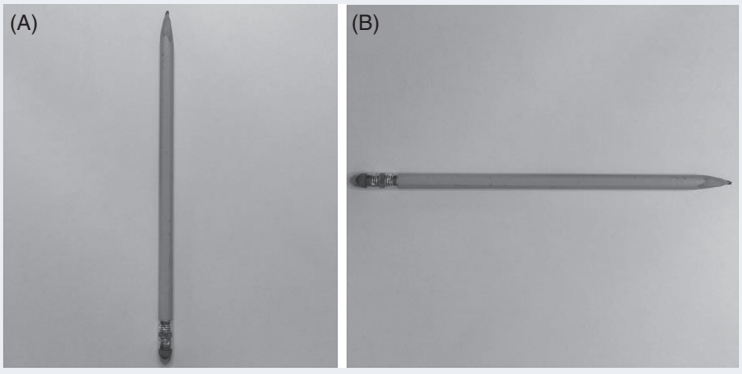
hand-action verbs. Rather than merely showing an association between language processes and motor processes, these studies reveal a *functional role* that motor processes play in language processes.

PERCEPTUAL SIMULATIONS

In addition to language input activating *motor* aspects of linguistic meaning, it can also activate *sensory* aspects of linguistic meaning. For example, when you read a sentence like "John put the pencil in the cup," part of how you understand the sentence isn't merely by activating your concepts of pencil and cup but also by generating a perceptual simulation (i.e., partial mental imagery) of what that event would look like (Barsalou, 1999). To test this, Stanfield and Zwaan (2001) showed participants sentences like that and then presented an image of an object after each one. In this sentence-picture verification task, participants were to indicate whether the object had been mentioned in the previous sentence. With the sentence, "John put the pencil in the cup," half of the participants were shown an image of a pencil oriented vertically and the other half were shown an image of the same pencil oriented horizontally (Figure 14.3). In both cases, the correct answer is "yes" to verify that the object was indeed mentioned in the sentence, but the participants who saw the vertical pencil responded faster (than those who saw the horizontal pencil) because its orientation matched their perceptual simulation of a pencil resting upright in a cup. Results like that would not happen with a nonembodied cognitive system that uses only amodal symbolic cognitive operations.

Recall the image schemas discussed in Chapter 9, where the meaning of a preposition like "over" has a schematic two-dimensional image associated with it, such as an abstract "X" subject following an arc over an abstract landmark square. These image schemas may be the building blocks of perceptual simulations. Thus, part of how one understands a sentence like "The ship sinks in the ocean" is by generating some partial mental imagery of an object traveling vertically downward. In much the same way that Shebani and Pulvermüller (2013) showed that motor activity can interfere with language about those same kinds of motor movements, Richardson et al. (2003) showed that language-induced image schemas can interfere with visual input in the same region of visual space. That is, if the central vertical region of mental visual space is partly involved in representing a sentence like "The ship sinks in the ocean," then trying to discriminate simple shapes in the upper and lower regions of the computer screen while the vertical region of mental visual space is instead devoting itself to a completely different task should disadvantage the visual processing of those simple shapes on the screen. Richardson et al. (2003) played audio clips of sentences like "The ship sinks in the ocean" or "The mechanic pulls the chain," while participants tried to discriminate briefly flashed squares and circles that would show up in the upper, lower, or side regions of the computer screen. While listening to a sentence that invoked a vertical image schema (because the verb was "sink" or "rise" etc.), participants were slightly slower at discriminating shapes in the upper or lower regions of the screen, compared to shapes on the left or right

Figure 14.3 After reading the sentence "John put the pencil in the cup," participants were faster to verify that a vertically oriented pencil (A) was an object mentioned in the sentence than they were to verify a horizontal pencil (B) (Stanfield & Zwaan, 2001).



sides of the screen. Conversely, while listening to a sentence that involved a horizontal image schema (because the verb was "pull" or "run," etc.), participants were slightly slower at discriminating shapes on the left or right sides of the screen, compared to shapes in the upper or lower regions of the screen (see also Bergen, Lindsay, Matlock, & Narayanan, 2007).

In fact, it even works the other way around. Not only can language input (and its perceptual simulations) interfere with a visual task, but visual input can interfere with a language task as well. Edmiston and Lupyan (2017) had participants verify properties of concepts (such as "Does a swan have a beak?" or "Is a donkey a mammal?") while a random visual animation of colorful flashing rectangles of varying size, color, and position was splashed onto the computer screen. Over headphones, participants would hear a concept property query, such as "Does it have a beak?" Then, at the very moment that the concept name was presented over headphones (e.g., "swan"), the flashing visual animation would play on the screen for a fraction of a second. Thus, just as the participant was generating a visual perceptual simulation of a swan, in order to verify that it has a beak, the visual input would flood their visual system with flashing colorful randomness. This visual disturbance interfered with participants' ability to verify *visual* properties of concepts (such as a swan's beak), significantly reducing accuracy. However, the visual disturbance did not interfere with nonvisual semantic/encyclopedic knowledge of

concepts, such as a donkey belonging to the mammal category. This result suggests that our conceptual knowledge is rooted, in part, in embodied modal sensory representations and not solely in amodal symbolic formats of information.

However, consider what happens when a symbolic logical operation must be applied to a perceptual simulation—for example, when a sentence proposition is negated, as in "The umbrella was not open." Generating a perceptual simulation of an open umbrella for that sentence would, of course, be an inaccurate representation of its meaning. But generating a perceptual simulation of a closed umbrella involves using a concept property (i.e., "closed") that is not mentioned in the sentence. Kaup, Lüdtke, and Zwaan (2006) found that, very shortly after a sentence like that, when a picture of an umbrella was presented on the screen, people could name it equally fast whether it was an open umbrella or closed one. However, when a 1.5-second delay was introduced between the end of the sentence and the presentation of the image, people were faster at naming the closed umbrella image than the open umbrella image. That 1.5-second delay allowed the correct perceptual simulation to eventually come to fruition (see also Anderson, Huette, Matlock, & Spivey, 2010). With binary properties, like open versus closed, a simple recurrent network model of sentence processing (that had sensorimotor features added) was able to simulate this process of using the negation word, *not*, before an adjective in order to reverse its meaning (Huette & Anderson, 2012). Even with a double negative, like "The shoe was not untied," the network settled on the correct meaning of the shoe being tied.

Although representing *logical negation* is somewhat tricky with perceptual simulations, it can be even more difficult to represent *abstract concepts* with perceptual simulations. What would *justice* even look like? Borghi and colleagues (2018) point out that learning concrete concepts, such as *table*, *dog*, and *umbrella*, is easy because the embodied sensory and motor representations are relatively consistent across experiences for those concepts. By contrast, the kinds of experiences in which you might learn what *justice* or *freedom* means are highly variable in their sensory and motor impressions. The one consistent source of information for learning about those abstract concepts is the social and linguistic context in which those events take place. Abstract concepts are acquired largely via conversations with other people about the meanings of those events. Therefore, Borghi, Flumini, Cimatti, Marocco, and Scorolli (2011) trained participants on a set of novel concepts, where some of them had concrete objects that they referred to, and others had abstract processes or events that they referred to. When prompted to verify properties of those concepts with a manual button press, participants were faster for the concrete concepts than for the abstract concepts. However, when they verified properties of those concepts with a vocal response into a microphone, they were faster with the abstract concepts than with the concrete concepts. A bit like the Tucker and Ellis (2001) findings with different hand affordances for precision grasp and whole-hand grasp, Borghi's participants had different response affordances for abstract and concrete concepts. Their knowledge of *concrete* concepts potentiated manual responses with the hand because that is how one would typically interface with examples of concrete concepts. By contrast, their knowledge of *abstract* concepts potentiated vocal responses

because talking with people is how one would typically interface with examples of abstract concepts.

Evaluating Embodied Cognition

What embodied cognition offers is an alternative to the traditional computer metaphor of the mind. Classical cognitive science treated cognition as relying solely on the software of rules and symbols that were no longer grounded in their original sensory and motor representations. In that computer metaphor of the mind, the frontal lobes were roughly assumed to be the cognitive system, using formal logical computation. The rest of the brain was roughly assumed to serve almost as the "sensorimotor circuit board" that is separate from cognition. That circuit board delivers sensory impressions to cognition, which then converts them into logical symbols for formal computation. The circuit board then patiently waits to receive movement commands from the cognitive system so the motor processes can produce actions. In this metaphor, cognition's internal operations are carried out completely independently of sensory and motor representations. By contrast, embodied cognition suggests that those sensory and motor impressions are themselves what cognition performs its operations on (Glenberg, Witt, & Metcalfe, 2013). In this perspective, cognitive processing is not independent from sensory and motor processing. Instead, they all interact dynamically.

Despite the compelling laboratory demonstrations of embodiment in cognition, there are still some challenges to this perceptual simulation account of cognition. For example, Goldinger, Papesh, Barnhart, Hansen, and Hout (2016) describe the theory of embodied cognition as little more than "a set of vague and flexible claims." Papesh (2015) reports a failure to replicate one of the commonly cited experiments that purports to support embodied cognition—not described here. Finally, Mahon and Caramazza (2008) suggest that cognitive embodiment effects are often subtle and therefore may not justify modifying traditional symbolic theories of cognition. They further suggest that some experimental findings that are interpreted as supporting embodiment can be explained as a kind of trivial "leakage" of information from concepts to sensory or motor processes. According to that view, the fact that people are faster to respond to an image of a vertical pencil (compared to a horizontal one) after reading "John put the pencil in the cup" (Stanfield & Zwaan, 2001) could be due simply to a one-way flow of activation from the symbolic representation of [INSIDE (CUP, PENCIL)] to a visual processor that simulates that image *outside of the cognitive system*. Thus, cognition itself would still rely solely on amodal symbolic computations in that scenario and would not be influenced by sensory processes. Although it is possible that some embodiment findings can be reinterpreted in this fashion, Shapiro (2019) points out that other experimental findings do not lend themselves to that reinterpretation. Some embodiment findings reveal the influence of sensorimotor processes as a *functional aspect* of the cognitive process itself. For example, Pulvermüller et al.'s (2005) TMS experiment described above shows an influence of motor activation on a language task that happens *during* the language task—not trivially afterward.

DYNAMICAL SYSTEMS THEORY

Embodied cognition may simply be an unavoidable result of the nonmodular interactive nature of neural connectivity in cortex (recall the discussion of network science in Chapter 7). If sensory, cognitive, and motor brain regions are sufficiently interconnected that they cannot help but continuously share much of their information with each other via their massive bidirectional synaptic projections, then it should be no surprise that sensory and motor processes naturally become active even during mental events that might seem purely cognitive (Richardson & Chemero, 2014).

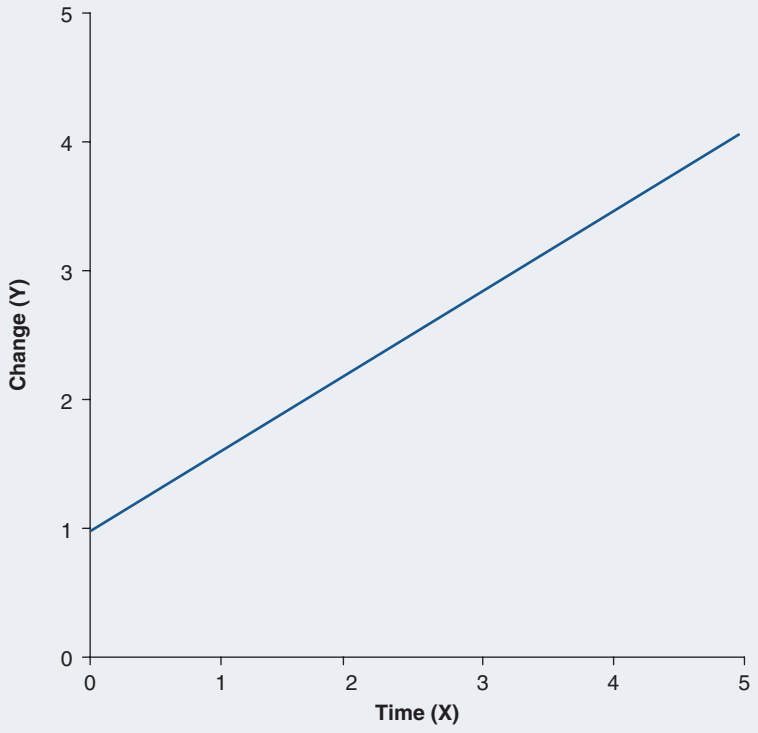
In this section, we explore some ways to study the complex dynamical system that comprises a network of networks like this. This perspective is called the **dynamical systems approach**, although sometimes it is referred to as chaos theory (Friedenberg, 2009) or complexity theory (Favela, 2020). We start off by describing the basic assumptions of the dynamical approach. These include the ideas of nonlinearity, predictability, state spaces, trajectories, and attractors. Once we have understood these, we are in a position to outline the major differences between classical cognitive science and the dynamical approach.

Nonlinearity

Nonlinear dynamical systems theory is the study of how systems change over time. A system in the most general sense is a collection of interacting parts. Change is a characteristic of all systems. The brain with its myriad interacting neurons is a good example of a complex changing system. We can plot change using a graph, with time on the *x*-axis and some measure of interest on the *y*-axis. Figure 14.4 shows a **linear relationship**. In this plot, the output is proportionate to the input. For every increase along the *x*-axis, there is a constant increase along the *y*-axis. There are instances of linear relationships in psychology, many that we have already studied. These include the mental rotation function, the serial visual search function, and the function depicting the scanning of items in working memory (from Chapters 4 and 5).

Nonlinear relationships are much more complex and difficult to predict. The output in these systems is not directly proportional to the input. A small change in input values can produce miniscule or dramatic changes in output. Figure 14.5 shows what a nonlinear change might look like. In this example, it is the logistic equation, which is used to study dynamical processes in physics, chemistry, biology, and also cognitive science. When the *x* value increases from –8 to –6, there is no appreciable effect on the *y* value. It just stays at zero. However, as the *x* value approaches zero, the *y* value begins to increase more and more rapidly. Then, as the *x* value approaches 4, 5, and 6, the effect on *y* becomes negligible again. The output of a dynamical system often serves as part of its own next input, which often leads to nonlinearity in that output. When our brain issues a command to move our hand to grasp an object, the perception we receive of our initial movement will feed back and alter our subsequent movements. In this way, we can fine-tune our movements to grasp at the precise location of the object. As a result, the velocity of that reaching movement changes nonlinearly over time. Nonlinearity is

Figure 14.4 A linear relationship of the kind we have seen throughout much of this book.

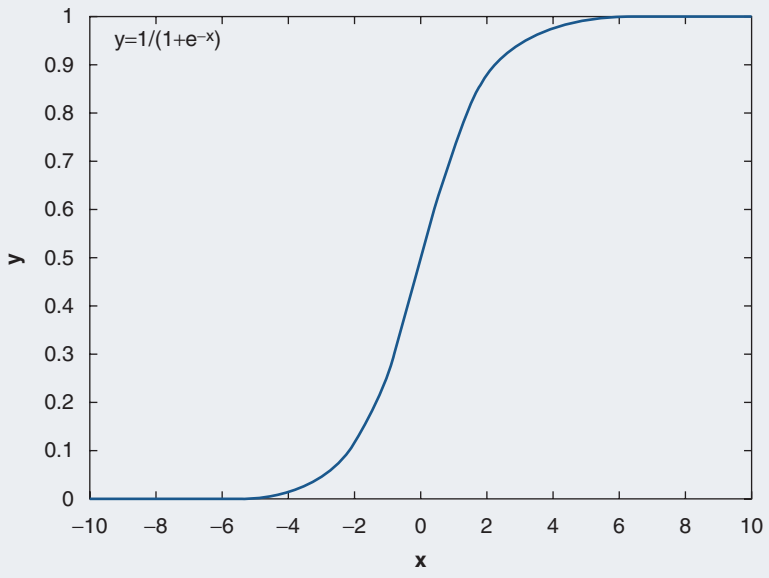


probably the norm for most psychological phenomena, and recent research provides us with numerous examples from psychophysics, perception, learning, memory, cognition, human development, motivation, and problem solving (Guastello, 2001).

Predictability

Chaotic systems such as the brain or mind can exhibit a wide range of possible behaviors. However, they are not random. A **random process** is one that is based strictly on chance and whose behavior can be described only statistically. We can say that a certain behavior has a particular probability of occurring in a chaotic system, but we can't say with certainty that it will. A **deterministic process** is one that is completely and, to within some degree of precision, exactly specified. Chaotic systems are deterministic but only predictable to a limited degree. We can, in many cases, write out the mathematical equation that describes how a chaotic system works, but without running this equation

Figure 14.5 The logistic equation is a well-studied example of a nonlinear relationship.



through a simulation with many time steps where some of its outputs become its next inputs, the equation alone will not tell us what the system will do in the future.

A defining feature of chaos is **sensitivity to initial conditions**. Two slightly different starting conditions in a chaotic system will eventually diverge to produce two very different outcomes. This idea is also referred to as the butterfly effect. In chaos theory, it has been suggested that a butterfly flapping its wings in Brazil could be the tipping point that causes a hurricane to hit Texas. The result of this is that in chaotic systems, we have short-term predictability only. We can extend the length of a prediction by increasing the precision of our starting values, but we can never be so precise that we obtain complete long-term predictability. This short-term predictability is at least better than random systems, where we have no predictability whatsoever. This means that chaotic behavior, although not as predictable as simple linear systems, is more predictable than random processes.

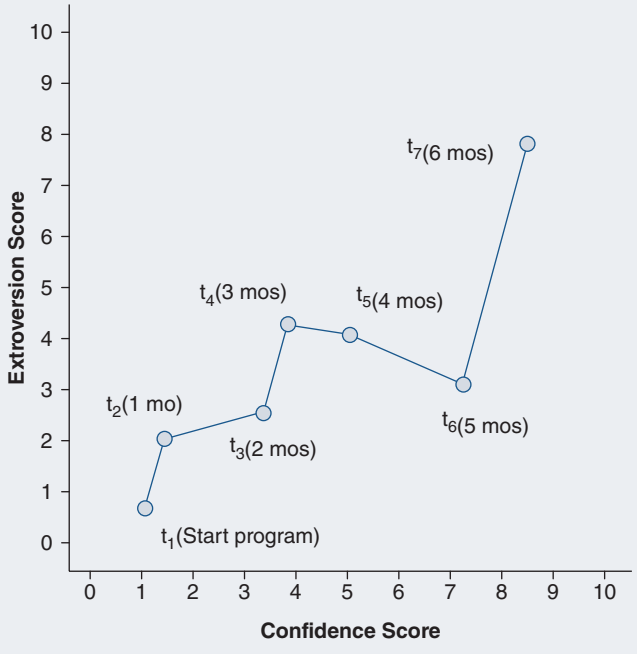
State Space and Trajectories

One way to measure the activity of a dynamic system is to plot how that activity changes over time. If we measure only one variable, we can see changes such as those depicted in

the previous figure. However, most systems change in more than one way. In this case, we need a way of capturing change for two or more variables simultaneously. We can do this using a **state space**, shown in Figure 14.6. Each axis of this graph represents a dimension for a single variable. A point within this space represents the state of a system at one point in time. A sequence of points shows how the system changes over time. This sequence is known as its **trajectory**. In Figure 14.6, the two axes represent different mood states. The line traveling through this space shows how both of these moods might change in a single person over the course of 6 months.

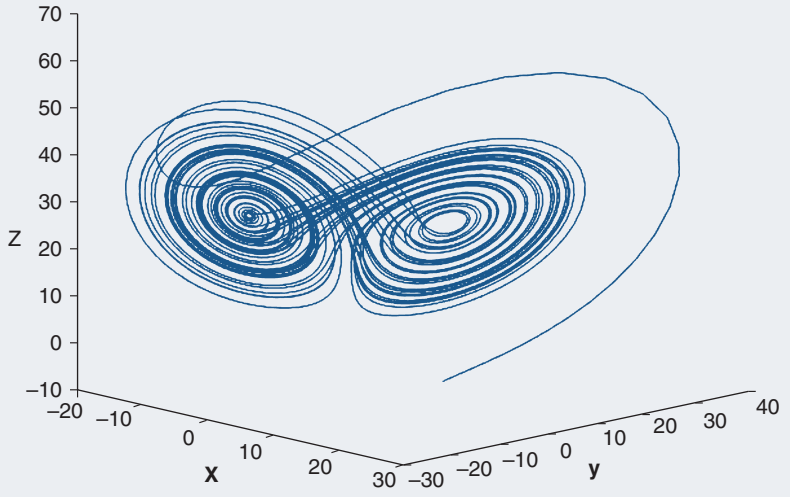
In the example shown here, our person starts off feeling low in confidence and introverted, perhaps because he hasn't met anyone at a new school. Eventually, though, he becomes more extroverted and confident. Maybe this is because he made several good friends. Trajectories can be complex, wandering all over their state space, but they can also be more ordered, in which case, they restrict themselves to one region of the space. Some ordered trajectories can also continually retrace a similar path through their space.

Figure 14.6 A psychological example of a state space with two dimensions.



Source: Friedenberg (2009).

Figure 14.7 The Lorenz attractor is a chaotic attractor that can start in any location in three-dimensional space and will always find its way to re-creating a version of this bent figure-eight shape as its three equations update one another over time.



Attractors

The behavior of a trajectory can be explained by the notion of an **attractor** in the state space. An attractor is a location in the state space that pulls, or attracts, the system toward itself. Imagine that a valley in a landscape is the attractor and a ball is the state of the system. If the ball is just inside the edge of the valley, it will roll down the sides until it comes to rest at the bottom. In this case, we could say that the valley floor is the attractor, but the sides of the valley are an attractor basin. If the ball falls inside the basin, it is "captured" and will wind its way down to the bottom. However, the ball doesn't always stay there. Other forces can knock it out of the valley and get it moving again.

There are several types of attractor. A **point attractor** is a single fixed point in a state space. It represents those systems that are constant on the variables measured and have stopped changing. For example, a person's mood may not change very much when he or she is quietly reading a book. **Periodic attractors** are also referred to as limit cycles. These consist of trajectories that loop back on themselves, following the same

general path through the space over and over again. A person suffering from bipolar disorder may cycle back and forth between manic and depressive states. His or her behavior could thus be described as following a periodic attractor.

There are other more complicated attractors. If a trajectory is a combination of two or more limit cycles, it can follow a path best described as a doughnut. This three-dimensional shape is a torus, and trajectories that follow it are caught in a toroidal attractor. A chaotic or **strange attractor** follows an even more tortuous route. It can wander erratically over a wide region of its state space, tracing out complex three-dimensional regions. One of the more famous strange attractors is the Lorenz attractor (Figure 14.7). No matter where in *x*, *y*, *z* coordinates it starts out, this system of three equations will always move toward creating a version of this bent figure-eight shape.

DYNAMICAL REPRESENTATION

The classical view in cognitive science, one that we already have presented in some detail, is that mental representations are static and discrete. They can be represented by symbols, such as words or symbolic propositions. These representations are said to be intentional; they are "about" what they refer to in the outside world, what we called referents. In that traditional view, if you thought about a table, then somewhere in the mind there would be a static enduring symbol that would stand for "table." The dynamical systems perspective views symbols in a different light and challenges this long-held notion (Peschl, 1997). There are at least three reasons for this, which we will examine next.

According to Peschl (1997), a representation is always a subjective construction created by the observer. It is produced in a unique way by the sensory and nervous system of an individual organism. Because the nervous system is highly dynamic, the representation is in constant flux, depending on the organism's state of mind. For instance, the representation we have may differ depending on what mood we are in, new information we have learned, and countless other factors. This means that no two representations will be alike, either between individuals or even within a single individual at different times (see also Casasanto & Lupyan, 2015).

The second reason is empirical. Experiments and simulations fail to show a stable referential relationship between a representation and a referent. If all networks were purely feed-forward, sending the results of their computations only forward in the network, then representations based on sensory input might be preserved. However, feed-forward networks are the exception rather than the rule in most nervous systems. Most neural mechanisms are recurrent networks. Information in recurrent networks feeds information back to previously processed information. The result is a distortion of the original sensory input. Peschl (1997) also notes that the environmental signal is already distorted by the transduction of sensory receptors before it even gets to be processed more fully by "downstream" mechanisms.

A final reason for why a dynamical systems perspective poses a challenge to classical cognitive science is that there is an improved alternative account of representations. This

is the **constructivist framework**. In this perspective, representations are constructed by a physical nervous system and are always system relative. Constructionists advocate that the primary purpose of a representation is to generate a behavior, not to generate an understanding of the environment. This behavior is created in the service of the organism's functional adaptation to its environment. The purpose of a representation in this view is not to preserve some idealized aspect of the environment or to induce a conscious experience but simply to map stimulus inputs onto motor outputs.

Evolutionary approaches such as the field of artificial life already approach representation from this kind of constructivist perspective. The genetic content of organisms in artificial life simulations can be thought of as a form of representation, but what they code for is a behavior that either helps or fails to assist the animal in surviving or reproducing. This point of view is also seen in robotics. In a reactive architecture, representations exist as production rules (IF–THEN statements), or artificial neural network connections, that map a stimulus onto a behavior (Brooks, 1991; see also Beer & Williams, 2015). Robots with this architecture are surprisingly good at navigating through complex environments and performing other motor tasks. Knowledge in both of these views does not exist in isolated representations but in the entire structure and architecture of the organism as it relates to its environment. The knowledge is embodied.

What do dynamical representations of this sort consist of? Peschl (1997) makes several suggestions. They could be a particular pattern of activation in a neuron population. Alternatively, they could be the synaptic weights between neurons. This second option corresponds to representations in artificial neural networks (ANNs), where the matrix of weights depicts the strength of connections between nodes. Representations can also be seen as sequences of neural activation patterns. This is the view favored by the dynamical system school. The sequences are patterns of activity that unfold over time in a neural system and correspond to trajectories and attractors in a neural state space.

INTERDISCIPLINARY CROSSROADS: MULTIPLE APPROACHES TO LEVELS OF EXPLANATION IN COGNITIVE SCIENCE

Bechtel and Abrahamsen (2006) differentiate between phenomena and mechanism in cognitive science explanation. Each corresponds to a particular level of description. A phenomenal explanation is symbolic and primarily descriptive in nature. By symbolic, they mean using symbols and rules that act on those symbols. Descriptive means that the phenomenon can be described in terms of what it is like but is not explained in terms of more fundamental processes. Linguistic explanations fall under this category. In contrast, mechanistic explanations are **subsymbolic**. They operate on a level underneath the symbols. In cognitive science, these would be the fundamental units of a brain or network, such as neurons or nodes, along with

their connections and how they operate. Mechanistic explanations are also productive; they explain the processes that give rise to particular phenomena. ANNs are a good example of mechanistic explanation because activity in the nodes and links that form the network can produce cognitive processes such as pattern recognition or concept formation.

The learning of past tense for English verbs is a good illustration for each of these levels. Linguistic representation in the form of symbols and rules applied to symbols can very efficiently capture the process of forming a past tense. For regular verbs, this could be expressed as V → V + ed, where the V stands for a regular verb stem linked to its past-tense form by a general rule of adding the "ed." For irregular verbs, the general rule cannot be applied, and we would need to look up the past-tense form individually for each verb. This can be expressed linguistically as *run* → *ran, buy* → *bought,* and so on, where the particular pairs of representations are individually linked. A mechanistic explanation of past-tense formation would require an explanation of how the process actually takes place in the brain. Alternately, we could construct an ANN that would perform the function, as has been done (Plunkett & Marchman, 1991; Rumelhart & McClelland, 1987).

Unfortunately, linguistic and connectionist accounts of cognition through history have often been pitted against each other. On one side are the linguists and philosophers who stick to using the symbolic notation and favor the descriptive level. On the other side, we have the computationalists and neuroscientists who favor subsymbolic mechanistic explanations. Bechtel and Abrahamsen (2006) argue that these two approaches ought to be viewed as complementary rather than antagonistic. Each can be better at accounting for particular issues. For example, linguistic descriptions are better for addressing the issue of systematicity, or how we can use knowledge of language to infer other kinds of meanings. Connectionist descriptions are better at explaining certain empirical results such as graceful degradation. There are examples of researchers from both sides of this divide who have collaborated to develop new theories that blend elements of both approaches. These include optimality theory and distributed statistical approaches to language learning (Prince & Smolensky, 1997; Saffran et al., 1996).

In addition to linguistic expression and connectionist networks, several other levels of description have been used in cognitive science (Bechtel & Abrahamsen, 2006). One example comes from the field of mathematical psychology that specifies the quantitative relations between variables. Well-known instances are Weber's law and Stevens's power law from psychophysics. Mathematical equations such as these, if taken in isolation, are not mechanistic, but they can be combined with other models to form a mechanistic account. Sternberg's (1966) working memory search model is an example of a mathematical model combined with an information-processing model (what we have been calling process models in this book). This is because it combines an equation specifying what the reaction time should be with a sequential processing model of the mental symbols corresponding to the probe items, target items, and matching procedures.

Dynamical systems theory, which we are describing in some detail in this chapter, corresponds to yet another level of description. If the dynamical approach sticks to mathematical formulations and the ways in which a system can change from one state to another, then it can be classified as a purely phenomenal level of description (Kelso, 1995). However, there are examples of blended approaches where dynamical systems theory has been combined with subsymbolic mechanisms (Elman, 2009; Tabor, 2009). The dynamical systems approach also has been successful in describing long-term developmental processes, and in some of this work, connectionist mechanisms have been included (Spencer, Thomas, & McClelland, 2009; Thelen & Smith, 1994).

DYNAMICAL VERSUS CLASSICAL COGNITIVE SCIENCE

The Continuity of Mind

In *The Continuity of Mind*, Spivey (2007) makes the case for adopting the nonlinear dynamical systems approach (see Table 14.1). He argues against the traditional notion of representations being categorical and discrete. The computer metaphor of mind is seen as no longer supported by contemporary research findings. It is suggested that cognitive science should discard the view of representations as symbolic and localized, of mental functioning as modular, and of serial computation.

In place of a sequence of symbols, the continuity of mind proposes instead that all mental activity is fuzzy, graded, and probabilistic. Thoughts are best characterized as trajectories in a state space. The dimensions of this space can correspond to activity levels in large neuron populations. Thought will sometimes gravitate toward attractor basins in this space that represent a percept or concept, but typically, they will pass through these regions rapidly. In the continuity of mind, the emphasis is on trajectory movement, not on the locations in state space in which they may dwell only briefly. In this section, we outline this continuity approach, contrasting five traditional conceptions of mind from classical cognitive science with the dynamical concepts that ought to replace them.

Table 14.1 A comparison of the classical and dynamical views in cognitive science (after Spivey, 2007).

Classical Cognitive Science	Dynamical Cognitive Science
Modularity: Brain consists of many independent special purpose modules.	*Distribularity:* Brain is only partially modular. Much of it consists of functionally interdependent subsystems.
Component dominant: Modules process information serially and only stop when they reach a stable state.	*Interaction dominant:* Computational units act in parallel and continuously share information with each other. System never reaches a stable state.
Internalism: Brain is a closed system that is mostly insulated from the surrounding body and world. Computations are relatively unaffected by ongoing changes in the environment.	*Externalism:* The brain is contextualized by the body, which is contextualized by the world. Interactions and information continuously flow back and forth between these different levels.
Amodal: Representations are primarily abstract symbols and need not be tied directly to their real-world sensory stimuli. The system does not need to act in "real time."	*Modal:* Cognitive systems are embodied are embedded in the world and interact with it in the here and now. Representations are tied directly to the sensory modalities from which they come.
Feed-forward: Information flow in the mind is primarily one-way. Stimuli trigger mental activity, which then triggers actions. There is no feedback where processes that occur later in that sequence influence processes earlier in it.	*Recurrent:* Information flow can be both forward and backward. Feedback loops and cyclical computations are assumed. As a result, all representations are extended in time.

Modularity Versus Distribularity

Recall our discussion of modularity, introduced in the first cognitive approach chapter. If the brain consists of many of these special-purpose mechanisms, then studying the mind becomes a bit easier. That is because we can take out or dissociate some modules and study the others in isolation. Removing or controlling one module in this kind of system allows us to see the role it plays, much like a missing piece in a jigsaw puzzle where all the pieces fit into one another in only one way. Because most of the important processing occurs inside a module, the interactions between them are simple, minimizing the difficulty of constructing a "causal map" or process model of the system's functions.

If the brain is not modular, then the traditional approach runs into problems. Explaining the brain now means having to account for countless complex interactions between processing centers. This certainly is the case if we reduce these mechanisms to the individual neuron level, since neuron firing rates are modulated by excitatory and inhibitory inputs from thousands of neighboring neurons. Now, we have a much more difficult time diagramming the system. Each module in some sense has now lost its "boundary": Its functions are local but also global, being tied to the operations of processing centers both near and far.

So is the brain modular or not? It is perhaps most accurate to say that it is partially modular. Portions of neural tissue are anatomically and morphologically distinct from one another and exhibit some degree of functional independence. By the same token, we also see many areas that are highly interdependent, such that the neural mechanisms that generate a particular cognitive process may be distributed across multiple disparate brain regions. As a way to refer to this blend of somewhat-distributed processing and somewhat-modular processing, Kingsbury and Finlay (2001) coined the term **distribularity**.

Component-Dominant Versus Interaction-Dominant Dynamics

If the brain were modular in the sense that Fodor (1983) means, it would display component-dominant dynamics. In this scenario, a cognitive module waits until it receives sufficient input or has performed enough localized computations to reach a stable state. Only then will it generate an output that is passed on to another module for continued processing. In a system with component-dominant dynamics, most of the important cognitive computations takes place *inside* each component. The interactions between those components perform only very simple computations. In a sense, the module can be thought of as being "selfish" in that it is keeping information to itself for as long as possible. This type of dynamic is seen in serial systems where computations are performed in a sequential order.

In contrast, the brain appears to exhibit interaction-dominant dynamics. Here, functional units continually update their neighbors. Subsystem A, for example, could share its partial results of processing with Subsystem B even before it has finished its own computations. McClelland (1979) referred to this as "processes in cascade." A system like this may never reach a stable state if its various cascaded subsystems also send feedback to one other. In a system with interaction-dominant dynamics, most of the important cognitive computations actually emerge among the interactions *between* the subsystems, not so much *inside* them. These subsystems can be thought of as "altruistic" because they

share their information with one another. This dynamic is more characteristic of distributed systems that use parallel distributed processing techniques.

The evidence supporting the interaction-dominant view is now overwhelming. Although classical box-and-arrow models of visual perception and language comprehension tended to assume a feedforward sequence of modules that carried out progressively more complex information processing, recent evidence soundly disproves this perspective. For example, neuroimaging techniques have demonstrated bidirectional functional anatomical linkages between disparate brain regions in language processing (Matsumoto et al., 2004) and also in visual processing (Gandhi, Heeger, & Boynton, 1999). Brain regions that are roughly specialized for processing the semantic meanings of words have been shown to send feedback to a brain region that is roughly specialized for processing individual phonemes (Gow & Olson, 2016). Brain regions associated with reasoning and decision making have been shown to send feedback to brain regions associated with visual object recognition (Kveraga, Boshyan, & Bar, 2007). We also see interactions between sensory modalities. Crosstalk between features takes place not just within a modality such as vision but also between modalities such as vision and touch. The visual system interacts with the touch system during tactile discrimination of orientation. We know this because the visual cortex becomes active during this task (Sathian, Zangaladze, Hoffman, & Grafton, 1997). Disruption of the visual cortex by transcranial magnetic stimulation also interferes with the task (Zangaladze, Epstein, Grafton, & Sathian, 1999).

Further evidence for interaction-dominant dynamics comes from looking at the time series of cognitive activity over the time scale of hours. When a system, such as a richly interconnected brain, carries out mostly interaction-dominant dynamics, one can expect that it will exhibit a particular statistical signature in its time series of data. By contrast, a component-dominant system exhibits a rather different statistical signature. When a feedforward modular model of cognitive processing is designed to produce variation in its behavior and even occasional errors (like humans do), it is done by adding white noise to the system's processing algorithms. The statistical signature for this white noise is a relatively steady and constant range of random variation over time. However, human cognitive behavior does not exhibit steady white noise in its variation over time. Instead, human behaviors tend to exhibit a statistical signature of "pink noise" in their variation over time, where the range of random variation oscillates up and down over time (Gilden, 2001; Kello, Anderson, Holden, & Van Orden, 2008). This pink noise statistical signature can only reasonably come from a system that has interaction-dominant dynamics (Van Orden et al., 2003).

Internalism Versus Externalism

Closed systems are insulated from what is occurring around them. In other words, they are spatially and functionally bounded. It seems fairly certain that the brain is an open system at almost every level of analysis (Yoshimi, 2012). On a small scale, neurons are influenced by their immediate neighbors as well as by larger neural circuits that participate in more widespread brain activity. The brain as open system is not an entirely new idea. Both the ecological movement in perception and the reactive architecture in robotics adopt this perspective. This debate about whether mind is just the brain or something

more has coalesced into two differing philosophies. **Internalism** holds that mind can be entirely explained by a description of the brain and of brain states (Segal, 2000). In opposition to this is **externalism**, advocating that mind is the result of interactions between the brain, the body, and the immediate physical environment (Clark & Chalmers, 1998).

Amodal Versus Modal Representations

According to Mahon (2015), a modular cognitive system that functions mostly independently from the sensory and motor systems to which it is attached is able use amodal symbols in its computations that are abstracted from the original sensory and motor modalities (e.g., vision, hearing, manual action). Those formal logical computations over amodal symbols can be performed in a manner that is unaffected by fine-grain details in the original sensory signal (such as changes in vocal pitch during speech perception or metric distances in visual imagery). The amodal symbols represent (or serve as internal proxies for) the objects and events in the world to which they refer, but they need not bear any resemblance to those objects and events.

By contrast, an interactive nonmodular cognitive system may perform its cognitive computations on the actual sensory and motor impressions that are generated in those sensory and motor brain regions themselves (i.e., patterns of neural activation in those modal brain regions that are roughly specialized for vision, hearing, or manual action). By performing its operations on modal representations, instead of amodal ones, this cognitive system retains the fine-grain details of the original sensory signal (such as changes in vocal pitch during speech recognition or metric distances in visual imagery). Many of the findings in embodied cognition detailed earlier in this chapter are best accommodated by a theory of cognition that uses modal representations rather than amodal representations (Barsalou, 2016; Shapiro, 2019).

Feed-Forward Versus Recurrent Pathways

One last point to be made concerns the direction of interactions between the brain, body, and world. According to the recurrent view, the causality here is bidirectional. The world acts on the body through afferent perceptual inputs, while the body acts on the world through efferent motor actions. Then, those motor actions directly influence what aspects of the world act on the body next. Thus, your actions cause changes in your sensory inputs just as much as those sensory inputs cause changes in your actions. According to the continuity of mind, it is the continuous loop or cycle between these that constitutes mind, not any of the one-way paths. It is thus best to think of mind as a **recurrent pathway**, one that loops back on itself, instead of as a **feed-forward pathway** that goes in only one direction.

Examples of two-way information paths are evident within the brain as well. For many years, it was believed that the visual system was a feed-forward system. The optic nerve projected information about an image from the retina to a relay center in the thalamus. From there, the signal was sent to cells in the primary visual cortex (Area V1). Following this, the input splits into temporal and parietal regions, constituting the ventral and dorsal pathways, respectively. It turns out that this is actually a very simplified

view. The visual system is now known to have synaptic projections backward as well as forward. For example, numerous pathways run from Area V1 back to the thalamus. The image sent from the eye is, indeed, passed upward but then also triggers downward cascading signals as part of the normal process of visual perception.

In a cascaded recurrent system, such as a brain with feedback loops, it can be difficult to know when one representation stops and the next begins in the temporal sequence. They may blend into one another over time as the system pursues a continuous trajectory through its state space (Spivey, 2007). The state of the system may resemble one representation for a brief while, but when it moves in state space to resemble the next representation, that transition may be so smooth and gradual that determining where to draw the line between the previous representation and the next one may become arbitrary. Without reliable discrete distinctions between one representation and the next, it would be difficult to identify amodal symbols on which to carry out logical computations.

Evaluating the Dynamical Perspective

Our tour of the dynamical approach has covered a lot of ground. We need to pause for a minute now and examine how well it addresses the issues facing cognitive science. Like the reactive approach in robotics, it sees cognition not only as "in the head" but also as a looping cyclical process between brain, body, and world. It also offers up a more accurate account of the way real-world variables interact. Linear relationships in nature are the exception, not the rule. The really interesting cognitive phenomena are most certainly nonlinear and change in a complex way over time. Although we may not be able to predict the behavior of complex systems such as the brain in the long run, we can at least conceptualize them more accurately.

The dynamical approach offers us a unifying methodology. Any field in science can be understood as the movement of a trajectory through state space. Whereas traditional cognitive psychology experiments evaluate one or two dependent variables, state spaces can contain any number of dimensions and allow us to see how they covary together. This multidimensional approach may allow us to better understand individual and cultural differences because we can follow the course of their system trajectories simultaneously across several critical dependent measures. We can also manipulate multiple independent variables and see their effect on trajectories and attractors. Perhaps even consciousness can be better comprehended using these techniques.

The mind is dynamic, not static. It is in a constant state of flux. In fact, it is questionable whether we ever have the same thought twice or experience exactly the same emotion twice. Many of the process models we looked at in this book are a useful first attempt at surveying mental processes, but they are oversimplified and fail to account for the complex interactions that occur between brain areas and neurons. A number of dynamical techniques are now available that allow us to look at changing processes better. These include oscillatory models based on cyclical neuron activity and differential-equation, cellular-automata, chaotic, and agent-based models (Friedenberg, 2009).

Dynamical cognitive science supports one side of several theoretical issues we have considered. It advocates a quasi-modular view of mind, where modules retain some

specialized processing but are also open to feedback and alteration by one another. Processing is thus interaction dominant instead of component dominant and involves both feed-forward and feedback pathways. Finally, human cognition is seen as both situated in the environment and embodied by its sensors and effectors.

The dynamic approach offers us an alternate take on representations as well. Representations in this account are not objective in the sense of having discrete static values. They are inherently subjective and constantly changing based on other cognitive states. Sensory input may shape the initial form of a representation, but this may be substantially modified with subsequent computations. The primary purpose of a representation is to aid in the execution of adaptive behavior, not to preserve an isomorphic relation between it and its referent.

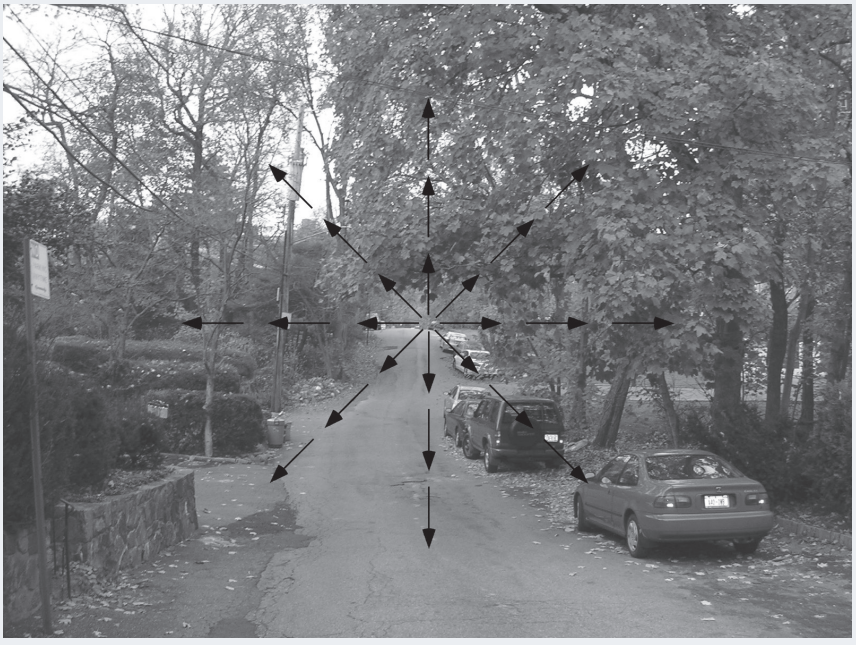
ECOLOGICAL AND EXTENDED COGNITION

Ecological Perception

Perception is a good example of where the representational–computational and sensory–motor views contrast. In cognitive science, the dominant view of perception has been that it is computational. All the theories summarized in the first cognitive chapter involved a complex series of stages where features were extracted and used to reconstruct an internal representation of the visual input. This process is inferential and indirect because the cognitive system must resolve uncertainties in that visual information through a time-consuming sequence of serial computations.

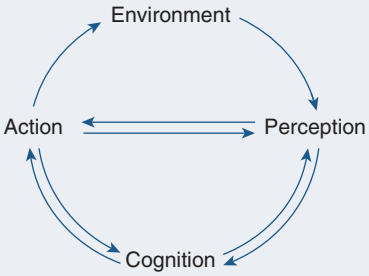
James J. Gibson (1979) proposed, instead, that perception involves the direct mapping of sensory inputs onto motor outputs, with little need for internal representations—much like the reactive architecture for robots in Chapter 13. He argued that perception results from the interplay between a moving observer and its environment. In this **ecological perception** view, the visual system uses perceptual information to directly carry out motor actions without resorting to mental representations and a series of computations. In turn, those motor actions make changes to the perceptual input that resolve uncertainties in the information. For instance, the ability to judge our direction of movement in space can come from optic flow, whereby surfaces move in the visual field in a coherent fashion. As you travel down the road in Figure 14.8, most regions of the image gradually slide leftward, rightward, upward, or downward. But the one spot that doesn't slide toward the periphery, the one spot that stays where it is and merely expands in image size, is the point of expansion. That is your heading. That is where you'll end up if you continue this direction of movement. Thus, without having to recognize any objects or compute any representations, this optic flow nonetheless provides enough information to know where your actions are taking you. Similarly, our ability to judge a distant object's size can be informed by the ratio of how much of that object is above the horizon and how much is below the horizon: the horizon ratio. Properties such as optic flow and the horizon ratio are examples of invariant sources of information that are directly available to the perceptual system. They do not require inferences or extended computation.

Figure 14.8 **Forward locomotion produces a global optical expansion of texture in the visual field, indicated here by these arrows. Information about speed is provided by the rate of expansion. The central point of expansion is your heading.**



In this ecological perspective, mental activity does not occur solely inside the brain, isolated from the surrounding environment. Much of our thinking is directly connected to sensory inputs and motor outputs. It is devoted to interacting with the "outside," as opposed to operating only on complex forms of representation and computation generated from the "inside." This idea is known as **extended cognition**. Figure 14.9 shows cognition as just one aspect of the mental activity that arises when an organism interacts with its environment. External stimuli that are processed by perception can lead to action directly, as when we reflexively catch a ball that is coming at us. They also affect it indirectly through cognition, as when we deliberately make a decision to throw a ball. Neural feedback loops allow prepared actions to influence cognition and perception, as well as for cognition to influence perception. Moreover, as actions change the environment, this in turn alters our perceptions, which in turn alter our cognition and action, and so on. For instance, when you lightly toss a ball straight

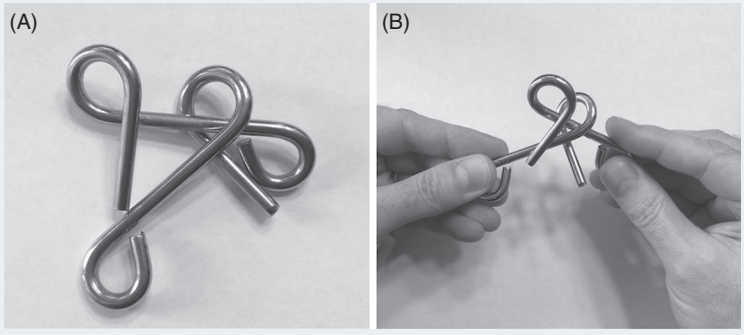
Figure 14.9 The perception–action cycle from which mental activity emerges.



up in the air, the force of that action determines what you perceive, as you watch it rise and then fall back into your hand. This perception–action cycle (Neisser, 1976) stands in sharp contrast to the classical feedforward view of the senses, according to which they just passively take in information with the sole purpose of matching it to a stored template in memory.

Ecological perception, for some tasks at least, seems to do away with the need for mental representations that intervene between perception and action (Stepp, Chemero, & Turvey, 2011). For example, Dreyfus (1992) contends that intelligence—and, consequently, the performance of efficacious action—does not require formal symbolic representation. Many cognitive operations can come about as a result of the interaction between an agent and its environment. For instance, imagine trying to solve a metal puzzle, where you must remove one metal ring shape from its tangle with the other metal shape(s). In principle, you could look at the metal puzzle, generate a mental representation of it, and perform mental rotation and other mental manipulations on that representation until you find the solution—all before even touching the metal puzzle (Figure 14.10A). But, of course, that is not what you actually do. You physically grasp the puzzle with your hands and perform those rotations and other manipulations on the object itself (Figure 14.10B). Many of the cognitive operations that lead to finding a solution for that metal puzzle are performed with the hands in the environment—not solely inside the brain. Simple learning principles based on such sensorimotor experiences are enough to allow for sophisticated interaction with a complex environment. We have seen this idea already. Brooks's (1991) reactive architecture, discussed in the robotics chapter, is based on the idea of reflexive responses to environmental stimuli. The robots constructed with this architecture have been quite successful in locomoting about the world without any representations at all.

Figure 14.10 **A metal wire puzzle could be cognitively processed via an internal mental representation (A) or, from an ecological perception perspective, it could be cognitively processed via physical interaction (B).**



Sensorimotor Interaction

There are many ways in which the interaction between an organism and its environment can generate cognitive processes over and above those going on inside the organism itself (Favela & Chemero, 2015; Turvey & Carello, 2012). For example, earthworms—which have no central nervous system—will line their burrow with leaf fragments that fit the size of the burrow, and they will even use small pebbles when appropriately sized leaf fragments are not available (Sims & Gerard, 1985). While it is unlikely that the earthworm has a "mental model" of what it wants its nest to look like, the affordances provided by the objects in its environment allow it to arrange that well-formed nest, nonetheless.

Primates with exquisite nervous systems also get in tune with the objects in their environment. Maravita and Iriki (2004) discovered neurons in the parietal cortex of monkeys that respond both to touch of the hand and also to visual input of stimuli near that hand. It is as though these multisensory neurons are coding for something like "graspability" of objects on or near that hand. Maravita and Iriki trained these monkeys with a small rake tool, for reaching a food object that is far away and dragging it closer. Before training, the visual receptive field for those multisensory neurons was rather small, extending just a few inches beyond the edge of the hand itself. If a food object was placed outside of that range, the multisensory neuron would not become excited. However, after training with the small rake tool, the visual receptive field for those neurons had expanded to include the full length of the rake tool. Thus, when a food object was placed far from the hand but near the rake that was in that hand, the neuron would produce a robust train of action potentials. Essentially, after learning how to use the rake,

the monkey's brain had grown to treat that rake (i.e., an external nonbiological object in the environment) as part of its own body schema.

This kind of phenomenon works with humans as well. Instead of inserting electrodes into the parietal cortex, Jessica Witt and Denny Proffitt (2008) had people estimate the distances of briefly flashed spots of light on a table while they were either holding a conductor's baton or not. After substantial practice with using the conductor's baton to touch distant locations, their verbal estimates of the distance of the spots of light were compressed by about 10% when the baton was in their hand, compared to when it was not in their hand. Basically, their practice with the conductor's baton had recalibrated their sense of how "reachable" things are, so that when it was in their hand, the spots of light on the table were perceived as slightly closer than they actually were. Just like with the monkeys, the conductor's baton had become part of their body schema and thus had altered their perception of affordances.

An external object can become part of your cognition when you use it in more continuous extended activities as well. Dotov, Nie, and Chemero (2010) had participants use a computer mouse to play a video game that involved herding an agent into a circular boundary (a little like herding sheep). As people developed skill at this game, their hand movements with the computer mouse revealed a substantial amount of "pink noise" in their variation over time. (Recall the discussion of pink noise in the section above on interaction-dominant dynamics.) The statistical presence of pink noise in the time series of hand movements suggests that the human-and-computer-mouse was behaving like *a single system*. However, the experimenters then introduced a brief disruption in the relationship between mouse movements and mouse-cursor movements on the computer screen. Instead of the mouse-cursor moving where the mouse-and-hand told it to, some random variation was added to the mouse-cursor's actual movement. Essentially, the tool (i.e., computer mouse and cursor), with which the person had become one, was now malfunctioning. And thus, the human-tool system was now malfunctioning. As a result, the presence of pink noise reduced significantly in the time series of human hand movements on that mouse. The malfunctioning of the mouse cursor caused that *human-tool system* to devolve back into two separate systems: a human and a tool.

Not only can the cognitive union of human and tool get disrupted, but it can also grow over time, with extended practice. Ramesh Balasubramaniam (2013) had participants practice for days at balancing a two-foot-long pole on the end of their finger for as long as possible. He recorded the position of the base of the pole in three-dimensional space over time and also the position of the person's center of mass in three-dimensional space over time. Early on, these two different time series were only slightly coordinated with one another. However, as skill improved, and people were balancing the pole for minutes at a time instead of several seconds, the coordination increased between position of the base of the pole and the position of the body's center of mass. When Balasubramaniam analyzed not just the pole's time series and not just the center-of-mass time series, but the combination of those two time series together, he found that skilled pole balancers were statistically best described as *a single system* comprising both pole and balancer. The two had become one *human-tool system*.

Extended Cognition

Similar to the brainless earthworm lining its burrow, a slime mold can also make its environment "do some of its thinking for it." A slime mold is a unicellular creature, with no multicellular nervous system at all. And yet it is able to "remember" locations where it has already been (so as to not waste time foraging for food again in the same place) because it can detect the slime trail that it leaves behind everywhere it goes (Reid, Latty, Dussutour, & Beekman, 2012). When you "jig" your environment with clues and reminders, such as a slime trail, the environment itself holds onto crucial information in relevant locations (Kirsh, 1995). When a group of agents leaves traces and tools in the environment as a result of their actions, and these traces and tools assist with future actions, it is called **stigmergy**. Humans do this kind of thing as well. For example, we might place our car keys on a hook by the front door. That way, instead of us having to remember where they are all the time, our environment simply shows us our keys perceptually, right when we need them the most.

Using another video game as an experiment, Kirsh and Maglio (1994) examined how expert Tetris players acquire their magnificent skill at the game. Rather than developing unusually impressive skills at mental rotation of internal representations of shapes, what expert Tetris players actually do is tap the "Rotate" button far more frequently than typical Tetris players do. Thus, instead of relying on neural cognitive processes such as the mental rotation discussed in Chapter 5, these experts were relying on their environment to perform that rotation of the object externally. By using their perceptual matching neural processes, which are much faster than anyone's mental rotation neural processes, these experts offloaded the difficult rotation task onto the computer interface in front of them. Thus, by pressing the "Rotate" button with their finger, they were using their hands (instead of their brain) to do some of their cognition for them—a little bit like the metal wire puzzle in Figure 14.10.

In addition to your hands, your eyes can do some of your cognition for you as well. Grant and Spivey (2003) recorded people's eye movements while they looked at a diagram of an infamous puzzle: Karl Duncker's (1945) tumor problem. What makes this problem so infamous is the fact that most people trying to solve it eventually give up or run out of time. It goes like this: Imagine that you are a doctor dealing with a patient who has an inoperable stomach tumor. You have gamma rays that destroy organic tissue at sufficient intensity. How can you cure the patient with these gamma rays and, at the same time, completely avoid harming the healthy tissue that surrounds the tumor? While trying to solve the problem, people generate a variety of imperfect solutions, such as carving an incision in the stomach lining to aim the gamma ray at the tumor. However, that would damage the healthy tissue. The solution to this problem, which usually arrives as a "sudden insight," requires having multiple weak gamma rays converge on the tumor from different directions. Each weak gamma ray travels harmlessly through the healthy tissue and then they meet at the tumor, where their combined radiation is sufficient to destroy the tumor. In her first experiment, Grant showed people a simple diagram with a disk in the middle (the tumor), surrounded by an oval (the stomach lining), and recorded their eye movements while they guessed at possible

solutions for a maximum of 10 minutes. When the one third of participants who solved the problem were 30 seconds away from finding the solution, the eye-tracking data showed that they were focusing especially on the stomach-lining oval. By contrast, the two thirds of participants who were 30 seconds away from running out of time did not pay any special attention to the stomach lining. If looking at that part of the diagram somehow helped induce the cognitive insight to have the gamma rays *converge* on the tumor, could the display be augmented in a manner to attract people's eyes to that part of the diagram and thus induce insight in people who otherwise might have failed? Her second experiment did just that, by animating the stomach-lining oval, making it blink one pixel thicker three times per second. This unavoidably drew people's eyes to the stomach lining and doubled the percentage of people solving Duncker's elusive tumor problem. Thus, a simple perceptual trick drew people's attention to a component of the diagram that helped them gain cognitive insight into a rather abstract concept of convergence to solve this difficult problem (see also Thomas & Lleras, 2007). According to the extended cognition perspective, their environment was doing some of their cognition for them.

Importantly, there is more in your environment that can become part of your cognition than just objects and events. There are also other people whose extended cognition may overlap with yours. Consider when someone hesitates near the end of sentence and you find yourself completing the sentence for them. Consider when you and a friend can't quite remember the name of a movie star, but when you say the first name out loud, your friend says the last name for you. A good conversation is much more than two people taking turns at sending messages to each other. A good conversation is a process of two speakers sharing and co-creating the dialog together. It requires resonance, anticipation, and coordination. When two people are engaged in a conversation about a shared task, the tiny bit of sway in both of their postures gets correlated in time (Shockley, Santana, & Fowler, 2003), the eye movements they make to objects in a shared display get correlated in time (Richardson, Dale, & Kirkham, 2007), and even their gestures and word choices get correlated in time (Louwerse, Dale, Bard, & Jeuniaux, 2012). Their whole bodies get coordinated with one another. They behave as *one system*.

Since motor movements are largely driven by neural systems, the only way for motor movements to get correlated between two people is if the neural activity patterns are also correlated between those two people. Kuhlen, Allefeld, and Haynes (2012) tested this idea. She videorecorded a woman telling a story and, at the same time, recorded her brain activity with EEG. She then did the same thing with a man telling a different story. Then, she superimposed the two videos on top of each other, producing a ghostly image of the two talking heads and an overlay of both audio tracks from the two stories. The video clip can be slightly distracting, but it is relatively easy to pay attention to one of the speakers and ignore the other. Kuhlen then recorded the brain activity of people viewing the video clip. Half of them were instructed to pay attention to the woman and her story, while the other half were instructed to pay attention to the man and his story. Participants who focused on the woman speaker tended to produce brain activity that was statistically correlated with her brain activity (and less correlated with the brain

activity of the male speaker). The reverse was true for listeners who focused on the male speaker. Essentially, listening to someone tell you a story makes your brain do some of the same kinds of things that the speaker's brain was doing.

Not only does action and neural activity get coordinated between two people during a shared task, but perception does as well. Fusaroli and colleagues (2012) examined data from a shared perception task, where participants carried out a difficult visual discrimination task and then discussed their *individual* responses to reach a consensus for their *shared* response to the visual stimulus. Occasionally, their individual responses conflicted, so the shared response would require that one of them change their answer. Fusaroli found that some of the conversation transcripts revealed very little correlation in the types of phrases they each used to express levels of confidence in their judgment. In those pairs of people, the shared responses were not as accurate as the responses of the best individual responder in the pair. Two heads were *not* better than one. For example, a highly accurate responder may give in too frequently to a less accurate responder, and thus their shared responses would be *less accurate* than the highly accurate responder's individual performance. However, Fusaroli found that when a pair of participants developed a strong correlation in the phrases they used to express levels of confidence (e.g., consistently using "positive," "pretty sure," and "not sure"), their shared responses outperformed that of their most accurate individual responder. Essentially, the pair was developing their own miniature language that allowed them to calibrate each other's level of confidence in their perception of a particular visual stimulus. This linguistic coordination made their two brains act like one system, with better performance than either brain alone. Sometimes two heads really are better than one—especially when they are both "on the same page" about how to discuss their perceptions.

Evaluating Ecological and Extended Cognition

Although the evidence in favor of ecological perception (Turvey, 2018) and extended cognition (Clark, 2008) has been accumulating in recent years, there remains a contingent of cognitive and neural scientists who continue to defend a crisp delineation between mental activity in the nervous system and informational influences (that do not constitute mental activity) in the environment. After all, while cognition may appear to be *embodied* in most cases, even to some degree with abstract concepts (Borghi et al., 2018), cognition surely is not always *extended*. It is rather easy to imagine examples where your mind is doing something mentally complex even while you are not interacting with the environment at all. Most people can do some math in their head without having to use their fingers or a pencil. While they may rely on some visuospatial imagery (thus showing that the task is embodied), they need not rely on externalizing those cognitive processes. Therefore, it may be wrong to claim that cognition is *always* extended.

Some philosophers of mind would argue that even the experimental demonstrations above do not constitute cognitive operations taking place outside the skull. Adams and Aizawa (2009) argue that tools and other people serve as useful props to crucially influence the cognition going on inside a person's brain, but their actions do

not constitute cognitive processes that belong to the person in question. They may even be described as forming a system together, but only the brains in that system are doing anything that can fairly be defined as cognitive (Segal, 2000). This criticism is sometimes used to make an exaggerated caricature of the claims of extended cognition, in which the tool is considered capable of thinking by itself. This is perhaps unfair. Obviously, extended cognition does not make the claim that a pencil resting on a table can carry out some cognition by itself, any more than its critics would claim that a frontal lobe resting on a table could carry out some cognition by itself. According to ecological and extended cognition, cognitive processes emerge from the interaction of these various parts, some of them being brain regions, some of them being body parts, and some of them being tools (Clark, 2008; Clark & Chalmers, 1998; Shapiro, 2019).

INTEGRATING COGNITIVE SCIENCE

We have seen throughout this book that each cognitive science approach by itself substantially advances our understanding of mind. We have also seen that an integrative approach in which the disciplines work collaboratively on a common problem can be an even more powerful tool for investigating mind. Some of the best insights into cognition emerge from the interaction of these various disciplines. We hope to have conveyed this through the "Interdisciplinary Crossroads" sections in each chapter. However, cross-disciplinary research isn't easy. Each of the approaches varies in the way it describes the mind and in the tools it uses to study it. In this section, we propose four kinds of integration from which cognitive science research could further benefit. These are integrations across disciplines, across levels of description, across methodologies, and across cultural differences.

Integration Across Disciplines

Most cognitive scientists are specialists. They are trained in the knowledge and tools of their own discipline but lack sufficient understanding of other allied disciplines. An artificial intelligence researcher may be adept at devising search algorithms to locate information in a database but not be familiar with the literature on human memory search, an area of expertise for a cognitive psychologist. Similarly, a neuroscientist may know which brain areas underlie memory retrieval but not be able to implement a neural network model of the process, which is a skill possessed by someone working within the network approach.

One of the unfortunate results of this is that the field of cognitive science has not yet succeeded at developing a coherent set of shared theoretical assumptions. The multidisciplinary approach, while making important advances, has yet to give us definitions of basic terms. For instance, what is mental representation? What is mental computation? The classical information-processing view of cognitive science believes that representation is symbolic and infers that only select types of transformations on these symbols may occur. The connectionists and neural network researchers see mental representation

in a different light. To them, a representation is a coherent pattern of activation across a population of nodes, and computations result from the spreading of activity through the distributed connection strengths in the network. Dynamical systems theorists may treat a representation as an approximate segment of an otherwise continuous trajectory through state space. Further still, some ecological perception researchers are comfortable with doing away with the very concept of "representation" entirely.

The lack of a single unifying theoretical framework in cognitive science is seen by some as a weakness. In his book *Consilience: The Unity of Knowledge*, E. O. Wilson (1999) echoes a sentiment felt by many scientists. He believes that all the sciences should be unified within a single framework. However, not even the most fundamental science of them all, physics, has been able to develop a unified field theory that accommodates both Einsteinian relativity and quantum mechanics. Of course, cognitive science, as an academic field of study, is quite a bit younger than physics, biology, or chemistry. Moreover, in some circumstances, a unifying theoretical framework that is too dogmatic can be a hindrance. When a discipline has become deeply invested in a particular theoretical framework, experimental demonstrations that violate the predictions of that theory may go unfairly dismissed for decades (Winfree, 1984). Nonetheless, the lack of a set of theoretical assumptions shared universally among those who call themselves cognitive scientists surely detracts somewhat from the ability to coordinate a common educational experience on the topic. The young field of cognitive science still has some ways to go in order to evolve from a collection of eager *multi*disciplinary endeavors into one mature coherent *inter*disciplinary field (Nuñéz et al., 2019).

There are two solutions to this problem. The first is to train individual researchers in the interdisciplinary perspective. Scientists who possess an understanding of cognitive science fields outside their own area may gain key insights into the mind that they would not ordinarily have had. This solution has its limitations, however. Scientific practice requires a high degree of specificity and focus, and researchers cannot be expected to acquire more than rudimentary training in multiple alternative disciplines. The most that can be expected is some degree of familiarity with the literature and techniques of disciplines outside one's own specific area.

A second solution mentioned in the introductory chapter is interdisciplinary cooperation. Here, researchers from different disciplines work together on a common project. For instance, a team made up of investigators from different disciplines might focus their research efforts on a single topic, such as grasping. Each investigator could then contribute the unique strengths of his or her discipline's theoretical insights and methodology. Recently formed interdisciplinary cognitive science departments and research centers sometimes employ philosophers, cognitive psychologists, linguists, computer scientists, and neuroscientists. Studies at these departments and centers have yielded scientific advances that were unlikely to be obtained in departments devoted to just one of those disciplines.

Integration Across Levels of Description

Recall, from Chapter 1, David Marr's three levels of description that make up the trilevel hypothesis: (1) the implementation level, (2) the algorithmic level, and (3) the computational

level. These levels in the order given correspond roughly to three basic questions about a cognitive process: (1) Where is it taking place? (2) How is it taking place? (3) Why is it taking place?

It is rare in cognitive science to find a theory that provides an explanation at all three levels. Rather, different approaches tend to emphasize one level over the others. Neuroscience and robotics tend to focus on the implementation level, since they either study the brain's existing architecture or design new mechanical and electronic architectures. The cognitive and artificial intelligence approaches, because of their models and programs specifying information flow, tend to emphasize the algorithmic level. The philosophy and evolutionary approaches tend to focus on the computational level, since they provide more abstract explanations for the purpose of particular cognitive processes.

What we need are more theories and models that integrate across two or more levels at once. The levels can then mutually constrain one another, leading to better and more accurate theories. To illustrate, researchers interested in understanding working memory might first use brain imaging to determine which areas of the brain are active during the performance of different tasks. From this, cognitive process models could be derived. Eventually, a computer program that uses some aspects of the process model could be implemented and tested in a robot.

Integration Across Methodologies

Many of the disciplines employ distinct methods. Philosophers use deductive reasoning and thought experiments. Psychologists employ the scientific method. Cognitive psychologists, artificial intelligence researchers, and those adopting the network approach use modeling. Neuroscientists use a combination of case studies, lesion methods, electrophysiology, and brain imaging. Roboticists build and test machines in a real-world environment. Each of these techniques has its strengths and weaknesses. A computer simulation of a cognitive process may solve the problem of how to execute a cognitive task in the smallest number of processing steps. People who run models without taking experimental results into account, however, do so at their own risk, since nature may have evolved a different or better solution.

In the first chapter on cognition, we discussed several models of attentional selection—some favoring early selection, others favoring late selection. Different models were formulated over time, with models being modified to accommodate new experimental data. This is a good example of integrating modeling and experimentation across two disciplines. The models were modified in response to experimental results, and then those improved models made novel predictions that could be further tested in the lab, and so on.

Integration Across Cultural Differences

Not only are the disciplines that study cognition quite diverse, but the agents whose cognition is *getting studied* are also quite diverse. Different cultures exhibit different cognitive styles. Therefore, just as we acknowledge that different virtues are brought to the study of the mind by philosophers, neuroscientists, linguists, and so on, we also need to

acknowledge that subtly different accounts of cognition may be required for individualist cultures, collectivist cultures, nonindustrialized cultures, and so forth.

Even though cognitive science usually attempts to build models of cognition that could apply to any human, people are obviously different from one another. Individuals have unique and, in some cases, radically different ways of thinking. This begs many interesting questions: How is it that people come to think differently? Can we model the cognition of a differently abled person? Does a person who has never used electronic communication technology think differently than someone who was raised on it?

This phenomenon comes into view on a larger level when we start to compare societies, cultures, or countries. Of course, differences exist here as well. Segall, Campbell, and Herskovits (1966) found that Americans and Europeans were more prone than other populations to experience several visual illusions. They attributed this difference to the exposure of Europeans and Americans to "carpentered" environments that contain many rectilinear forms. There is some evidence to suggest that cultural experience plays a role in the perception of space (Deregowski, 1989). There are also personality differences among those living in different cultures. Nisbett (2004) has suggested that Asians think more holistically and see the "big picture," while Americans and western Europeans think more analytically, seeing the trees more than the forest. These studies suggest that perception and cognition are partly shaped by culture—as well as by language (recall the discussion of the linguistic relativity hypothesis in Chapter 9).

Conclusions derived from Western society psychology experiments may not be very representative of humanity as a whole. The most commonly studied human participants in cognitive science can be classified as WEIRD—Western, Educated, Industrialized, Rich, and Democratic. Henrich, Heine, and Norenzayan (2009) found substantial variability in experimental results across different world populations. The standard Western participants, when compared cross-culturally, were not near the middle of the distribution but were in fact outliers—their results were on the fringes of the samples tested. This was true for a wide domain of tests, including visual perception, spatial cognition, categorization, inferential induction, and economic decision making. Henrich et al. recommend that researchers must justify the generalizability of their results and compare their findings with those obtained from other cultures. Evolutionary psychologists play a crucial role here as they often test hypotheses in diverse populations. They also suggest that researchers should report and share as much of their data as possible to facilitate meta-analysis (statistical compilation of multiple studies).

The historical trend in cognitive science has been the study of universal characteristics—what all people or cognitive devices share in common. Research efforts have traditionally focused on solving problems that every information processor in a particular domain must solve, such as the size constancy problem in vision or the consolidation problem in memory. Cognitive science has devoted less attention to how or why these processes may be carried out differently in different individuals or cultures.

One way to study cultural differences is **anthropology**. Anthropologists investigate when and where humans first evolved and why they have varied with respect to some physical traits and cultural practices (Ember & Ember, 1985). Because of its focus, anthropology is a useful adjunct to the evolutionary perspective. This discipline can also tell us a lot about which thought processes are universal and which are particular to

specific cultures. Anthropology can help us elucidate the factors that account for such differences. Anthropologists who have made significant contributions to our understanding of mind include Lucien Levy-Bruhl, Franz Boas, Claude Levi-Strauss, and Dan Sperber (Gardner, 1985).

The quest to understand universal phenomena is certainly important and is arguably the first goal of an emerging science. A prime motivation for this quest is that it helps us understand the basic processes involved. The details can be worked out afterward. However, it is sometimes these very same details that can illuminate the operation of the more basic processes. For example, understanding why some people are so much better at math than others can probably tell us a lot about basic mathematical ability and the function of working memory. Research on universals and particulars ought to proceed in parallel. In this fashion, each can inform the other.

THE BENEFITS OF COGNITIVE SCIENCE

Cognitive science has made an indelible stamp on the way people now think about mind. Prior to its appearance, there was a plethora of different theoretical approaches to mind. One needs only to look at psychology to see this. Psychology, throughout most of the first half of the 20th century, generated a profusion of theories on what the mind is and how it should be studied. There was very little common ground on which researchers in different theoretical camps could stand. A psychoanalyst's conception of mental processes was qualitatively different from that of a Gestalt researcher. A behaviorist would have little to say to a structuralist.

The adoption of the cognitive view has meant progress in the area of bringing these diverse theoretical perspectives together—even if they do not always agree. The impact of cognitive theory can be seen in the ways it has influenced what were formerly considered separate disciplines. There are now cognitive theories of social behavior, personality, therapy, and education. Another indication of the integrating character of the cognitive approach can be seen in the ascent of concepts that are implemented across disciplines. The concept of a schema, for example, is important in both cognitive psychology and artificial intelligence.

Apart from theory, cognitive science has yielded a multitude of practical applications. Research in artificial intelligence has given us impressive programs that can recognize speech and diagnose complex medical and engineering problems. In the area of robotics, we see the development of new and more sophisticated robots capable of executing complex tasks, ranging from cleaning houses to disarming bombs. These applications have a long-term positive economic impact, as they allow businesses to operate more efficiently and can result in the creation of new industries. In the area of human–computer interaction, knowledge of human cognition, user experience, and software interface design has led to more intuitive computing interfaces that allow anyone to learn to use computing technology.

Practical results have come out of other cognitive science disciplines as well. Advances in the neurosciences often result in new treatments for disorders such as Parkinson's disease and Alzheimer's disease. The formulation of cognitive theories in

psychology has provided new therapies for the treatment of anxiety and depression. Cognitive insights also have had an impact on education, having led to new methods in the teaching of reading, writing, and other subjects.

THE FUTURE

We leave you with a few parting words. Cognitive science is unique in history because it groups such a large number of disciplines under one umbrella. This provides an unparalleled opportunity for collaboration. Cooperation and competition among the various approaches are both necessary for the advancement of our understanding of mind. We can only hope that more researchers follow this path.

But what does the future hold for cognitive science? Where are we going from here? Scientific progress relies on a number of factors. Among these are theoretical breakthroughs that reorganize our understanding. Technological advances also play a prominent role. Cognitive science, perhaps more than other disciplines, depends on new tools for measuring and simulating mental processes. Recent decades have seen the development of many new tools. These include optogenetics, magnetoencephalography, and the knife-edge scanning microscope. It is likely that other new tools like these are on the way. It is not unrealistic to imagine brain imaging devices that will be able to record individual neuronal activity of large cell populations or instruments that may be able to measure quantum fluctuations at the atomic level. The use of such tools will open a whole new vista of mental territory fertile for exploration. The future for cognitive science is bright indeed.

SUMMING UP: A REVIEW OF CHAPTER 14

1. According to embodied cognition, cognitive processes are not independent from perceptual and motor processes. Cognitive processes recruit perceptual and motor simulations in order to carry out their computations.
2. The dynamical systems approach may help explain what happens when cognitive, perceptual, and motor processes all interact with each other in real time. In this approach, there is an emphasis on nonlinear relationships, chaotic behavior, and how complex systems change over time.
3. According to the dynamical perspective, representations are subjective, constructed, and constantly undergoing change. The purpose of a representation is to guide behavior, not to form a discrete and stable symbol that stands for its referent.
4. There are a number of differences between the classical view in cognitive science and the dynamical systems approach. The dynamical view sees mental activity as fuzzy, graded, and probabilistic. Modules are only partially independent; cognitive processing is interaction

dominant and in constant interaction with the environment. Cognition is embodied and extended, and pathways of information flow are recurrent as well as feed-forward.

5. The ecological perception approach suggests that sensation can lead directly to action, without the need to stop at an intermediate cognitive relay station and compute internal representations. This sentiment is echoed in the reactive architecture in robotics.
6. The extended cognition approach suggests that, while brains certainly perform cognitive operations, the dynamic interdependence between brains, bodies, tools, and even other people encourages one to treat them all as one integrated system from which cognition emerges.
7. If we are to truly unite cognitive science, then integration across levels of description, across disciplines, across methodologies, and across cultural differences must take place. Bringing together experts in the various disciplines to work collaboratively on these topics may be the most successful approach for ensuring that cognitive science continues to grow into a mature coherent field of study.
8. Cognitive science has produced many benefits to society, including advances in education, engineering, and medicine.
9. The development of new technologies such as magnetoencephalography and knife-edge scanning microscopes will further bolster the development of cognitive science.

SUGGESTED READINGS

Chemero, A. (2009). *Radical embodied cognitive science*. Cambridge: MIT Press.

Friedenberg, J. (2009). *Dynamical psychology: Complexity, self-organization, and mind*. Litchfield Park, AZ: Emergent.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston, MA: Houghton-Mifflin.

Spivey, M. J. (2020). *Who you are: The science of connectedness.* Cambridge: MIT Press.

Turvey, M. T. (2018). *Lectures on perception: An ecological perspective*. New York, NY: Routledge.

Varela, F. J., Thompson, E., & Rosch, E. (2017). *The embodied mind, revised edition: Cognitive science and human experience*. Cambridge: MIT Press.

Wilson, E. O. (1999). *Consilience: The unity of knowledge.* New York, NY: Random House.

GLOSSARY

2½-D sketch An image representation that includes information about surfaces and layout

3-D sketch A three-dimensional image representation in which object parts are linked together by axes of symmetry and elongation

accommodate When we change our internal concepts or schemas when taking new information into account

acoustic code A memory code based on the sounds of the items

action potential The electrical signal that results when the inputs to a neuron exceed its threshold of excitation

action schemas In the Norman and Shallice model of attention, units that govern the execution of context-specific actions

adaptation The process by which an animal species changes in response to its environment through variation, inheritance, and selection

affect infusion model A model developed to explain the conditions under which emotions and moods infuse (influence) cognitive processing

affective computing Computing that relates to, arises from, or deliberately influences emotions and other affective phenomena

affordance a relation between an organism and an object (not inside either of them) that refers to a potential action that the organism could carry out on that object

agraphia A deficit in writing

alexia A deficit in reading

algorithm A formal procedure or system that acts on informational representations

algorithmic level A level of information processing that specifies exactly how the data are to be transformed, including what the steps are and what the order of the steps is

analogy A mental representation used to compare known information with new information to which it is similar

analytical rumination hypothesis The view that depression is an evolved response to complex problems. The depressive state induces a focus on the problem while at the same time minimizing distractions that might detour attention, working memory, and other processing resources.

angular gyrus A brain area located posterior to Wernicke's area. Damage to this area can produce problems in reading and writing.

anterior A direction indicating "toward the front"

anterograde amnesia An inability to remember new information after a traumatic incident

anthropodenial A blindness to the humanlike characteristics of animals or the animal-like characteristics of ourselves

anthropology The discipline that studies cultural differences by investigating when and where humans first appeared on the earth and why they have varied in some physical traits. Anthropology is also interested in how and why societies vary in their customary ideas and practices.

anthropomorphism The tendency to ascribe humanlike characteristics when explaining animal behavior

aphasia A neurological language deficit

apperceptive agnosia A type of agnosia in which people cannot assemble the pieces or features of an object into a meaningful whole

appropriate causal relation The relation between sensory inputs, motor outputs, and the representations that come between them

arbitrary The property by which there need be no relationship between linguistic symbols and their referents. Any symbol could be used to stand for any concept.

arcuate fasciculus A pathway that connects Broca's and Wernicke's areas

aristocratic network A type of small-world network where long-distance connections between nodes are the result of hubs

arousal The extent of physiological activation reflected in measures such as cardiac and respiratory rates

articulatory loop A system in the working memory model where speech- and sound-related information is rehearsed

artificial intelligence (AI) Concerned with the computational aspects of an intelligent agent

artificial life (Alife) The study of human-made systems that behave in ways characteristic of natural living systems

artificial neural network (ANN) A computer simulation of how populations of real neurons might perform some task

assertives A type of speech in which the speaker asserts his or her belief

assimilation When information is understood and interpreted in terms of what we already know

associative agnosia A type of agnosia in which people perceive a whole object but have difficulty assigning a name or label to it

attention Concentrated mental activity in which processing resources are allocated to information sources

attention-deficit hyperactivity disorder (ADHD) A psychological disorder characterized by distractibility, impulsivity, and mood swings. Patients have difficulty planning or executing plans.

attitude A learned predisposition to respond to a particular object in a particular way

attractors A region of a state space where the system is located or spends much of its time. They can be of different shapes.

attribution An explanation we give for the causes of events or behaviors

autism A neurodevelopmental disorder characterized by impaired social interaction and accompanied by restricted and repetitive behavior. These individuals are often aloof during childhood and remain egocentric as adults. They show impairments in verbal and nonverbal communication.

autobiographical memory Memory for personal events or episodes in one's life

automatic attentional processes Processes that do not require conscious attentional control

axon A long, tubular structure that projects outward from a neuron cell body and serves to transmit the action potential over distance

babbling stage A stage in language development where infants produce many sounds occurring in the languages of the world

backpropagation See "generalized delta rule or backpropagation"

basal ganglia A collection of brain areas involved in voluntary motor responses

base-rate fallacy A neglect of the probability that an item belongs to a particular category

basic emotions Those that are believed to be mostly innate and, thus, in humans ought to be universal or found in all cultures

behavioral economics A perspective that focuses on how evolution shaped the way we make decisions involving money

behaviorism A way of conceiving of empirical constraints on psychological states

belief-in-a-just-world phenomenon The phenomenon in which we think that people get what they deserve

binocular rivalry A phenomenon in which two perceptions alternately vie for visual consciousness

biological plausibility The idea that artificial neural networks effectively represent and model the characteristics of real-world brains

bipolar disorder A disorder in which individuals alternate back and forth between depressive and manic symptoms. The manic state is characterized by exuberance, a grandiose sense of self, decreased need for sleep, and excessive talkativeness.

bottleneck theories Theories that describe why it is that of all the information presented to us, only a small amount actually gets through to conscious awareness

brain imaging A technique that allows researchers to see the static three-dimensional organization of brain areas and the dynamic activity of these areas over time

Broca's aphasia A language deficit in which comprehension is intact but there are impairments in speech pronunciation and production

Cannon–Bard theory of emotion The theory that an event by itself can trigger an emotion. Although the event causes physiological arousal, the arousal is not necessary for the formation of the emotion. So in this view, the event immediately causes both the emotion and the bodily changes. The two do not depend on each other.

capacity theories Theories that conceptualize attention as a limited resource that must be spread around to different informational sources

case study A method in neuroscience in which researchers examine the relationship between brain damage and behavior in individuals

catastrophic interference This occurs when a network has learned to recognize an old set of patterns and then is called on to learn a new set. The learning of the new set modifies the weights of the network in such a way that the old set is forgotten.

category confusion effect A difficulty in differentiating between individual members of a given race

cell assembly According to Hebb, a small group of neurons that repeatedly stimulate themselves

central processing unit (CPU) That part of a computer that executes the instructions and controls the sequence of instructions to be carried out

centrality How a network coordinates information. This can be accomplished through a "leader" that receives information, evaluates it, and issues commands.

cephalization index A measure of brain size expressed as a proportion of body size that can be used to compare brain sizes between different animal species

cerebral hemispheres The two halves of the cortex, each with corresponding functional specializations. The left hemisphere is more analytic, serial, and logical. The right hemisphere is more synthetic, parallel, and relational.

chaotic systems Those who exhibit a wide range of behaviors. The behaviors can be ordered or disordered but not random.

cheater detection The ability to detect who has undeservedly received a benefit

Chinese room scenario A hypothetical situation in which a man uses a set of instructions to produce replies to questions in Chinese. It argues that the man can never learn to understand Chinese.

chunking The grouping of items together into a single meaningful whole in short-term memory

cingulate cortex A region of the cortex that is implicated in selecting a response, especially when that response requires inhibiting or ignoring another alternative response

classical conditioning A form of learning in which two repeatedly paired stimuli become associated so that one that formerly did not elicit a response now does

classical dualism The belief that both mental and physical realms are possible where the mind controls the body

clearness The clarity of a sensation determined by the amount of attention paid to it

closure The perceptual principle stating that shapes forming closed objects go together

clustering The probability that two nodes linked to a third will also be linked to each other. This network statistic reflects the interconnectedness of neighborhoods.

coarticulation Differential pronunciation of a given phoneme based on the phonemes that come before or after it

cocktail party effect The ability to automatically hear one's name spoken at a distance even though one's attention is focused on the immediate surroundings

coding In relation to computers, refers to program instructions

CogAff architecture A generic scheme for how cognitive–emotional processing may take place in either a person or a machine. It consists of the reactive, deliberative, and reflective stages.

cognitive architectures Specify the structure and function of many different cognitive systems and how they interact

cognitive dissonance A negative state that occurs when an attitude and a behavior or two attitudes are inconsistent

cognitive economy The principle that concepts should not have to be coded for more times than is necessary

cognitive map A mental representation that can explain learning without resorting to pure stimulus–response associations

cognitive neuroscience The study of the physiological structures and processes underlying cognitive function

cognitive psychology The study of human knowledge representation and use

cognitive science The scientific interdisciplinary study of the mind

cognitive theory of emotion The theory that it is the total situation and not just the arousal that determines emotions

commissives A type of speech that commits the speaker to some later action

communication The production, transmission, and comprehension of information in a language between individuals

comparative cognition An evolutionary perspective where researchers examine the similarities and differences in cognitive ability between different animal species

complex cells Cells in the visual system that code for an oriented line segment moving in a particular direction

complex ideas Ideas formed from the active mental combination of simple ideas

component-dominant dynamics A brain state whereby functional units must wait until they receive sufficient input or have performed enough local computations to reach a stable state

computation The transformation or manipulation of information

computational level An abstract level of analysis that asks what type of problem a computation solves and how it may have arisen

concept An idea that represents a class of entities that have been grouped together

conceptually driven process A process where context and higher-level knowledge aid in recognition

conditioned response In classical conditioning, a response elicited by a conditioned stimulus

conditioned stimulus In classical conditioning, a stimulus that elicits a response only after being repeatedly paired with an unconditioned stimulus

conduction aphasia Results from damage to the arcuate fasciculus; characterized by a difficulty in repeating words that have just been heard

conjunction fallacy A neglect of the conjunction rule, which states that the probability of simultaneously being a member of two categories is always less than the probability of being a member of either category alone

connectionism The study of mental operations through the construction and testing of artificial neural networks

conscious That aspect of mind that contains those thoughts and feelings of which we are aware and can directly access

consciousness The subjective quality of experience

consolidation The process by which information is transferred from working or short-term memory to long-term memory

constructivist framework A perspective in which representations are constructed by a physical nervous system and are always system relative. The primary purpose of a representation is to generate a behavior, not to depict the environment.

contention scheduling The scheduling that governs routine habitual performances

contralateral Refers to the opposite side of the brain or neural structure

control group The group in an experiment that does not receive the independent variable

controlled attentional processes Processes that require conscious attentional control

cooing stage The earliest period in language development, when infants exercise their vocal cords and mouths prior to articulation

corpus callosum A collection of fibers that connects the two cerebral hemispheres

counterbalance Systematically varying the order or type of conditions across experiment participants in an equally balanced fashion.

creative synthesis A principle by which the mind actively organizes mental elements together such that

the resulting whole contains new properties. These new properties cannot be explained by the characteristics of the individual elements themselves.

critical period A period of development during which linguistic experience is crucial for future language use. If a child is not exposed to language during this time, he or she may never acquire it or may suffer severe language impairments.

cultural intelligence hypothesis The view that complex social interaction created a form of social intelligence that then laid the groundwork for more general intelligence to emerge later

data-driven process A process that is driven entirely by the physical characteristics of a stimulus and fails to take into account the larger context or meaning

decay The loss of information over time from memory

declarative knowledge Knowledge that represents facts

declarative memory Memory for knowledge about facts and events. It is demonstrated by saying and occurs with conscious recall.

declaratives A type of speech where the utterance itself is the action

deductive reasoning The application of the rules of logic to statements about the world; if the premise statements are correct, then the concluding statement must be as well.

deep learning A type of artificial neural network that has multiple hidden layers, uses a version of backpropagation, and finds statistical patterns in very large datasets.

deep representations A type of image structure that consists of information in long-term memory that is used to generate a surface representation

defense mechanisms Constructs of the ego that reduce or redirect anxiety in various ways

degree of fan A concept that is related to many others has a high degree of fan. The node representing the concept will subsequently have many links radiating outward from it to other nodes.

deliberative Actions that are composed of previously developed behaviors that are integrated into a resultant behavior and may include external and internal stimuli

demand characteristics When participants in a study do what the stimuli, task, or experimental situation seems to demand of them

dependent variable A factor measured or observed by the experimenter to see if a change has taken place

depth perception A visual process by which the world is perceived three-dimensionally. It involves using cues from the environment as well as differences in processing between the two eyes to determine what is close and what is farther away.

designation The notion that a symbol expression can designate an object if it can affect the object itself or behave in ways that depend on the object

deterministic process One that is completely and to within some degree of precision exactly specified

dichotic listening task A task used to study selective attention. It requires a participant wearing headphones to listen to two different messages simultaneously played over each ear while paying attention to only one.

directives A type of speech in which a command is issued

disorganized thinking A symptom of schizophrenia characterized by a loose association between ideas. It can manifest as "word salad" in which concepts are not clearly related to each other.

displacement The ability of language to refer to something that is removed in space or time

distributarity A term to describe how the brain may be only partially modular, with different functions spread across different regions

distributed coding/representation A coding in which an object or feature is represented by a pattern of activation among a group of cells or nodes

divergent thinking A style of thinking where the goal is to generate many different ideas about a topic in a short period of time

divided attention A form of attention that can be split or divided among several alternative information sources

domain-general mechanisms Those that can operate on any type of information

domain-specific mechanisms Those that are tuned to perform special operations on only a certain type of information

dorsal A direction indicating "toward the top"

dorsal visual pathway A pathway that travels upward to the parietal lobe, where information about motion and location is extracted

dualism The belief that the mental and physical consist of different states or substances

duration How long a sensation persists

dynamic The characteristic of languages to constantly change as new words are added and grammatical rules altered

dynamical perspective A perspective that views the mind as constantly changing as it adapts to new information

dynamical systems approach A scientific interdisciplinary approach with the view that natural systems are complex and nonlinear. Activity in such systems is measured by using trajectories and attractors in state spaces.

early selection model Models of attention where information is selected early, based on physical stimulus characteristics

easy problems of consciousness Those problems that can be solved by cognitive science and that can be explained in terms of computational or neural mechanisms

echoic memory Auditory sensory memory

ecological models of cognition Models of cognitive processes that are acquired through learning

ecological niche The specific environment each animal species has adapted to

ecological perception The theory that perception is immediate and direct, not requiring representation or computation

egalitarian network A type of small-world network that contains mostly local short-distance connections with a few random long-distance connections

ego The aspect of mind that balances the competing demands of the id and superego; operates on the reality principle

electrical stimulation A method where an electrical current is passed through a bipolar electrode, causing the neurons in a localized area of brain tissue to become active. The resulting activity and behavior are observed.

electroencephalogram (EEG) A measure of the brain's gross electrical action

eliminativism The philosophical view that advocates abandoning words describing the mind and using only words describing the brain

emergence The idea that the features of a whole are not solely dependent on the features of the parts that make it up

emergence synthesis The theory that there are some emotions that are triggered solely by a stimulus and don't require any interpretation

emergent property A property that is realized through the interaction of a system's parts

emotion A relatively brief episode of coordinated brain, autonomic, and behavioral changes that facilitate a response to an external or internal event of significance for the organism

empiricism The view that knowledge is acquired through experience

encoding The name of the process by which information is taken in and converted into a usable mental form

endowment effect The tendency to stay with what we have and to prefer it over other options

enduring disposition Refers to automatic influences on where attention gets directed

engram A physical change in a specific location of the brain associated with learning

environment of evolutionary adaptation (EEA) The period of time during which many human psychological mechanisms are believed to have evolved

environmental dependency syndrome A symptom of executive dysfunction in which an environmental stimulus can automatically trigger an associated behavior

epilepsy A disorder in which neurons fire uncontrollably, producing muscle spasms and seizures

episodic memory A type of declarative memory containing knowledge of episodes or personally experienced events

epistemology The branch of philosophy devoted to the study of knowledge

equipotentiality The principle that many different parts of the brain seem to participate in memory storage

error signal In artificial neural networks, the difference between the actual and observed output. The error signal is used to alter connection strengths and train the network.

event-related potentials (ERPs) An EEG recording of brain activity in response to a particular event

evolutionary psychology (EP) A field of psychology that studies how evolutionary forces have shaped human mental capacities

evolved psychological mechanism A cognitive process that has evolved to solve a specific adaptive problem

exaptation (or neutral drift) Random mutations that produce new genes with little or no consequence for reproduction. They can spread in a population and at some point may assume a new adaptive function to a subsequent environmental change.

executive control system A system in the working memory model whose function is to initiate and control ongoing processes

executive dysfunction A disorder characterized by a broad number of deficits, including difficulty performing goal-directed behaviors as a result of frontal lobe damage

executive function Cognitive operations involved in planning, sequencing of behavior, flexible use of information, and goal attainment

exhaustive search When a memory search process continues matching all the way to the end of a list, even if the target already has been located

experimental group In an experiment, the group receiving the independent variable

experimental philosophy A form of philosophy that uses empirical methods, typically in the form of surveys that provide constructed scenarios. The surveys are used to assess people's intuitions in order to help answer philosophical questions.

explanatory gap The problem that subjective experience may not be fully explained by an objective account using physical and mechanical processes; the gulf between an objective and subjective description of mental phenomena

expressives A type of speech that describes the psychological states of the speaker

extended cognition The theory that some of the mechanisms that actually carry out cognition are not inside the organism but outside it (such as affordances, stigmergy, and certain handheld tools)

extensity The extent to which a sensation fills or occupies space

externalism The view that the mind is the result of interactions between the brain, body, and physical environment

fallacies A fundamental misunderstanding of a statistical rule that can result from applying a heuristic

feed-forward pathways Those in which information flow travels only in one direction

feel-good, do-good phenomenon A mood-enhancing experience such as finding money or recalling a happy event that can cause people to engage in helping behavior, such as picking up someone's papers or giving away money

feelings Correspond to the subjective experience of emotions. They are what it is like to feel afraid, sad, or jubilant.

field theory The idea in physics that objects are acted on by forces in fields

finite state model Contemporary implementation of Turing's universal computing machine

fissure A large spacing or separation between two areas of brain tissue

flashbulb memory Better recall of personal events during physiologically arousing or emergency situations

focused attention stage A later stage in pattern recognition that requires concentrated attention under voluntary control

folk psychology Theories of mind that use subjective terms and commonsense or intuitive reasoning

foot-in-the-door phenomenon A phenomenon in which people who agree to a small request now may comply with a larger one later

foraging The act of looking or searching for food

formal symbol manipulator A system that operates on contentless symbols based on a set of rules

formal logical system A system of symbols that can be combined into expressions that can then be manipulated by processes to form new expressions

framing effects In behavioral economics, the heuristic by which people prefer choices that are framed positively rather than negatively

free association What happens when one is presented with an idea and asked to produce whatever related ideas come to mind without censoring or inhibiting

frontal lobe Located on the anterior portion of the cerebral hemispheres. It is implicated in problem solving and motor behavior.

functional kinds Things that are distinguished by their actions or tendencies

functional magnetic resonance imaging (fMRI) A variation of MRI that is used to show changes in brain activity over time

functionalism The view that mental states are not just physical states but the functioning or operation of those physical states

fundamental attribution error The tendency to prefer internal or dispositional traits as the best explanation for people's behavior

fundamental utilities of consciousness The role consciousness plays in the survival of the organism

fusiform face area (FFA) Pictures of faces activate cells in this area, located in the temporal lobe

gambler's fallacy The belief that probability outcomes are not independent, that the probability of an event can be influenced by its past history

game theory Attempts to mathematically capture behavior in strategic situations, or games, in which an individual's success in making choices depends on the choices of others

general intelligence hypothesis The view that human intelligence did not evolve in response to particular selection forces such as social organization. Instead, it emerged in a general form that could be suited to solving all kinds of problems, whether they are social or not.

generalization The ability to apply a learned rule to a novel situation

generalized delta rule or backpropagation A way of training a network by repeatedly applying error feedback to alter connection strengths. The feedback is the difference between the actual and the desired output.

general-purpose processor The view that the mind can solve any type of problem equally well. It is based on the notion that a problem is solved by context-independent symbol representation and computation.

generative The property of language whereby symbolic elements can be combined to create a large number of meanings

geon A basic volumetric shape, such as a cube or a cylinder, that may be used in recognizing an object

gestalt The idea that a whole is more than just the sum of its parts

goal The desired end state or solution to a problem

graceful degradation A slow and gradual decrease in the performance of a network with increased damage

guided search A forceful or willed search for a memory item based on intelligence and reasoning

gyrus A ridge or fold of neural tissue

halo effect States that we tend to see positive traits group together. For example, if a person is good-looking, we tend to think he or she is successful.

hard problems of consciousness Those problems that require explaining the subjective quality of consciousness

Hebb rule States that if two connected neurons are both active simultaneously, the synapse between them will be strengthened

hemispatial neglect An attention disorder in which the patient ignores one half of his or her body as well as the surroundings on that side. Typically, this is for the left side and is the result of damage to the contralateral hemisphere.

heuristic This is a mental "rule of thumb" or strategy that acts as a fast and easy way of problem solving. Heuristics are right most, but not all, of the time.

hidden layer The second layer of a three-layer network. This is where the input layer sends its signals. It performs intermediary processing.

hierarchical networks Organized into different levels. The number of nodes decreases traveling upward with each group of nodes converging onto another group above it with fewer elements.

hierarchical organization A semantic network where concept nodes are arranged in different levels along a continuum from abstract to concrete

hierarchical paradigm A highly sequential robotic architecture in which planning is a key element of its design; also referred to as a "top-down" approach to robotic design

homunculus A hypothetical "little man" inside the mind who interprets and understands information

hubs Nodes that have many more links than the rest of the network's nodes

hypercomplex cells Cells in the visual system that code for angles (two conjoined oriented line segments) moving in a particular direction

hypothesis A statement derived from a theory that concerns potential outcomes in an experiment

iconic memory Visual sensory memory

id The aspect of mind that contains unconscious impulses and desires such as sex and hunger; operates on the pleasure principle

idealism The belief that only mental entities are real, that the universe is essentially nonphysical

idealistic principle Motivates the individual to do what it considers proper

illumination A flash of insight, a sort of "Aha!" experience where a solution comes suddenly to awareness

image generation Occurs when the encodings in long-term memory are used to form an image in the visual buffer

image inspection Occurs when an individual is asked some question about an image. It consists of a number of distinct operations such as "zoom," "pan," or "scan." These operations preserve the image while extracting information from it.

image transformation Refers to an operation performed on an image, such as a mental rotation. These operations alter the image.

immediate experience An individual's direct awareness of something

implementational level A "hardware" level of analysis that specifies the physical processes that carry out a computation

implicit association task An experimental task in which a person must judge whether a central word corresponds to one of two categories, each defined by peripheral words, using one of two buttons

impression The opinion or belief that you form about another person

incubation The period of time during which a problem is put aside. During this period, there are no conscious attempts at problem solving, but the unconscious mind may be attempting or have discovered a solution.

independent variable In an experiment, a factor manipulated or altered by the experimenter to see if it will cause a change

inductive reasoning A form of thinking in which commonalities about specific instances in the world are noticed and used to draw conclusions

ingroup favoritism The belief that ingroup members possess more positive traits in comparison with members of outgroups

inheritance The process by which parents pass on some of their genetic characteristics to their offspring

input layer The first layer of a three-layer network that receives stimulus input and where the stimulus is represented

insight learning The apparent spontaneous understanding of relationships that produces a solution to a problem

intensity The strength of a sensation

intentionality The relationship between a representation and what it stands for

interaction-dominant dynamics A brain state whereby functional units continually update their neighbors. Computation here is distributed over different areas that exchange information back and forth.

interference The inability of a network to distinguish similar patterns from one another

internalism The view that the mind can be entirely explained by a description of brain states

intervening variables Variables that mediate between initiating causes and a behavior

intraparietal sulcus Located in the parietal lobe, it may be responsible for the allocation of attentional resources and the binding of features in visual search

introspection The process of "looking inward" to experience and describe mental objects or states

ipsilateral Refers to the same side of the brain or neural structure

James–Lange theory of emotion The theory that an event produces arousal and other physiological changes in the body. These physiological changes are then interpreted. The result of the interpretation is the emotion.

joint attention The ability to coordinate attention with a social partner

knife-edge scanning microscope (KESM) A device capable of recording three-dimensional structures at the cellular level. It operates by slicing through tissue and reflecting an image to a camera.

knowledge-based approach A type of problem-solving method in which one conceptualizes the problem and its solution in terms of symbols and transformations on the symbols; representative of the traditional or classical approach in cognitive science

late selection model Models of attention where information is selected late, based on semantic stimulus characteristics

latent learning An animal's ability to acquire behaviors through experience without being subject to any reinforcement

lateral A direction indicating "toward the side"

lattices Regularly arranged matrices of parts used to demonstrate the principles of perceptual organization

learning When some event causes a change in the nervous system that in turn causes a change in behavior

lesion study A method in neuroscience where researchers deliberately destroy brain areas in animals and examine the resulting behavioral deficits

lexigrams Geometric patterns that can be used to create a language system

linear relationships Relationships between two variables where the output of one variable is proportional to the input of another

linguistic relativity hypothesis Also known as the Sapir–Whorf hypothesis. The strong version states that thought and language are so similar that it may be impossible to express the thoughts of one language in another. The weak version denies this but admits that the language a person speaks influences the way he or she thinks.

links The connection between each node in an artificial network

literal encodings Encodings that contain lists of coordinates detailing where to place points in the surface matrix to depict the represented object

lobes Cortical regions with different functional specializations

localist coding/representation Representation in the form of activation or activity in a single node in a network

long-term potentiation (LTP) The enduring facilitation of synaptic transmission that occurs following activation of a synapse by repeated intense high-frequency stimulation

loss aversion The principle by which people are more sensitive to losses than to equivalent gains

magnetoencephalography (MEG) A brain imaging technique in which small changes in magnetic fields are measured to determine brain activity. It has a spatial resolution of 1 millimeter and a temporal resolution of several milliseconds.

major depressive disorder Intense and long-lasting sadness that can occur without any precipitating incident. This can be accompanied by interruptions in sleep, appetite, and concentration.

masking When one stimulus immediately follows another and prevents processing of the first

maximal onset principle The principle that consonants usually precede vowels and typically form the onset of syllabic groupings; a feature found in all languages

maze A complex branching puzzle through which a solver must find a route

means–ends analysis A reasoning paradigm in which the intelligence notes the current state of a system and chooses some action that will reduce the difference between the current state and a specified goal state

medial A direction indicating "toward the middle"

mediate experiences Experiences that come from mental reflection about an object

memory The capacity to retain information over time; also the part of a computer that stores the instructions to be executed, partial computational results, and final results

mental accounting The principle by which we reason differently about money depending on the overall amounts. Generally, people are less likely to make efforts to save cash when the relative amount of money they are dealing with is large.

mental operations How a mental process operates, what it accomplishes, and under what conditions it occurs

metacognition Any process that monitors, regulates, or controls a cognitive process

metaphysics The branch of philosophy that examines the nature of reality

mirror neurons Neurons that fire when an animal performs some action and when the animal observes another animal performing the same action

modal memory model An early model of memory showing how information is transferred between sensory, short-term, and long-term memory

modularity of mind A theory that states that the mind is made up of innate and functionally distinct modules

molecular drive Occurs when a copy of a gene mutates and serves an adaptive function even though it was not selected for

momentary intention Refers to the conscious decision to pay attention to something

monism The belief that the mental and physical consist of a single type of state or substance

mood-congruent memory Increased memory for stimuli if the content of those stimuli matches a mood we were in while learning them

mood-dependent memory Increased memory for items when the mood at recall matches that during learning. Here the same mood must be present at both occasions.

moods Diffuse affective states that are often of lower intensity than emotion but considerably longer in duration

morphemes Units of spoken language that have meaning

morphological computation When certain perceptual and cognitive computations are performed by the shape or texture of the body parts of the agent, instead of being performed by its internal neural circuitry.

morphology The rules governing the word structure of a language

mosaic model of development A model in which brain mechanisms are almost entirely determined by genes and operate quickly, and the parts of the system develop independently

multimode model of attention The view that selection can be based on multiple stimulus characteristics such as the physical and semantic

multiple realization The view that any given mental state, such as pain, can be instantiated or realized differently in different creatures

multipleunit recording A brain recording technique where an electrode is used to measure the collective electrical activity of a group of neurons

nativism A belief that a significant body of knowledge is innate or "built into" an organism

nature–nurture debate A controversy that centers on the relative contribution of biology and experience in determining any particular capacity

network science The scientific interdisciplinary study of how complex networks operate

neural correlates of consciousness (NCC) The minimal set of neural events and structures sufficient for a specific conscious experience

neurons Cells that receive and transmit electrical impulses. They form the basic units of the nervous system.

neuroscience The study of nervous system anatomy and physiology

neurotransmitters Molecules released from the terminal button that travel across the synaptic cleft and attach to receptor sites on the surface of another cell

node A representation of a concept or element of the world

nonlinear relationships Relationships between two variables where the output of one variable is not directly proportionate to the input of another variable. A small change in input values can produce small or large changes in output.

object permanence The ability to understand that an object still continues to exist even though it is no longer perceptually evident

object-centered description An object representation that is described relative to the object itself and can be recognized from a variety of perspectives

obsessive–compulsive disorder A psychological disorder characterized by anxiety-induced thoughts of contamination and repetitive behaviors to reduce that anxiety

occipital lobe Located on the posterior portion of the cerebral hemispheres, it mediates visual processing

one-word stage A stage in language development where words are first used in a symbolic and semantic fashion

operant conditioning Using behavioral antecedents and their consequents to influence the occurrence and form of behavior; a form of learning in which reward and punishment shape the behavior of an entity (e.g., human)

operationalize To define the specific method of measurement for a dependent variable

optogenetics A method to measure and control neurons that have been genetically sensitized to light

ordered network All the connections are local. Nodes are connected only to other nodes in their immediate neighborhood.

ordinate A level of concept category organization of moderate specificity

outgroup homogeneity effect The belief that all people in an external group look alike

output layer The third layer of a three-layer network. This generates a representation of the response based on inputs from the hidden layer.

pandemonium model A feature detection model of pattern recognition in which "demons" are used to represent different aspects of a stimulus

panic attack A short-lived but intense episode of fear and sympathetic nervous system arousal

paradigm A philosophy or approach for developing theories for, analyzing, and evaluating a class of problems

parallel distributed processing (PDP) A type of information processing where large numbers of computing units perform their calculations simultaneously. Computational units can receive and process multiple inputs and transmit multiple outputs.

parallel memory search A memory search in which all the items in a list are matched against a target all at once

parallel search Visual search where all the items in the display can be matched for the target at once

parietal lobe Located dorsally, it is implicated in somatosensory, attention, and spatial processing

partial-report condition A condition in iconic memory studies in which participants are cued to remember only one row of letters in the display

pattern completion The perceptual process of filling in the blanks of an incomplete stimulus

pattern recognition The ability to identify objects in the environment

perception The process by which we gather and interpret information from the outside world via the senses; within a mechanical environment, perception refers to the process of scanning an environment and converting the resultant information into a set of abstractions comprising objects, features, and relationships

perceptional categorization deficit A type of apperceptive agnosia that involves difficulty recognizing objects when they are viewed from unusual angles or are lit unevenly

perceptron Artificial neural networks that detect and recognize information about the world, store this information, and use it in some fashion. They are characterized by the ability to learn from experience and can alter their connection strengths.

perceptual simulation A method of representing a concept by activating its perceptual components

percolating cluster A single giant group of susceptible sites connected by open bonds that occupies the majority of a network

percolation The propagation of a signal through a network. For example, the way in which a disease can travel through a social network, infecting other individuals.

periodic attractors (or limit cycles) Repetitive trajectories that loop or cycle back on themselves

phase sequence According to Hebb, a group of connected cell assemblies that fire together or close in time

phenomenal concept of mind The idea of mind as a conscious experience

phenomenology Refers to subjective experience rather than objective description

philosophical behaviorism The idea that mental states are dispositions or tendencies to behave in certain ways under certain circumstances

philosophy A discipline concerned with the search for wisdom and knowledge

phobia A persistent and irrational fear of a specific object or situation

phonaestheme A speech sound (often a consonant cluster) that tends to be correlated with a particular semantic feature.

phoneme The smallest unit of sound in a language

phonology The rules governing the sound system of a language

phrase structure The hierarchical organization of a sentence; represented using a tree diagram that shows the relationships between words

phrase structure grammar The grammar that governs the use of phrase structures

physical kinds Things that are identified by their material composition only

physical symbol system hypothesis (PSSH) The idea that a formal system can allow for intelligence even in computers

physicalism Also known as materialism; the belief that the universe is entirely physical

plasticity The ability of the brain to recover and reorganize following injury, primarily by rerouting or regrowth of damaged pathways

pleasure principle Motivates the immediate satisfaction of desires

point attractors Single fixed points in the state space

pop out A perceptual phenomenon that occurs when a target is easy to find because it differs from the surrounding distracters along a single stimulus dimension, such as color

positron emission tomography (PET) A brain imaging technique in which blood flow through the brain is measured while a participant carries out a cognitive task. This is accomplished using radioactive isotopes attached to tracer molecules such as glucose or oxygen. Brain areas that are more active will use more of the tracer molecules. This increase in resulting activity can then be detected.

posterior A direction indicating "toward the back"

posttraumatic stress disorder Anxiety and psychological reenactment in the form of flashbacks and nightmares caused by experience of a traumatic event

pragnanz Meaning "good figure," the Gestalt principle stating that parts that are simple will group together

preattentive stage An early stage in pattern recognition that happens automatically and effortlessly. It does not require conscious focused attention.

preconscious That aspect of mind that we can bring into awareness with effort

preferential attachment The principle by which nodes turn into hubs. Those with more links will gradually be linked to more often because they serve as good portals to gain access to large numbers of other nodes.

prejudice A decision or opinion we already have made

preparation An understanding of and preliminary attempts to solve a problem

primary motor cortex Located on the precentral gyrus, it contains a spatial representation or map of different body areas used to control motor behavior

primary somatosensory cortex Located on the postcentral gyrus, it contains a spatial representation or map of different body areas used to process sensory information from the body surface

priming Occurs when processing of a stimulus is facilitated by the prior presentation of a related stimulus

principles of perceptual organization Ways in which visual parts group together to form objects

prisoner's dilemma A social dilemma in which two prisoners must make decisions based on how the other is perceived to decide

proactive interference Occurs when information that is learned earlier interferes with remembering information learned later

problem solving A type of behavior that involves deciding that a problem exists and considering one or several solutions to the problem. Special-purpose solutions depend on the circumstances of a particular task.

General-purpose solutions can be applied to a broad variety of problems.

problem space The initial, intermediate, and goal states of the problem

procedural knowledge Knowledge that represents skills

procedural memory Memory for skill knowledge. It is demonstrated by doing and occurs without conscious recall.

procedure Algorithmic processes

process model A diagrammatic model that represents the way human information is processed. In the model, boxes are used to designate each stage or step in an information-processing sequence.

production rule (conditional statements) Propositions or predicates whose truth or falsity can be determined

property dualism The view that the mind and the body are made of the same stuff but have different properties

proposition A statement or assertion typically posed in the form of a simple sentence

propositional encodings Abstract, language-like representations, similar to declarative statements. They contain information about an object's parts, its location, and size.

prosopagnosia An inability to recognize faces, despite the capacity to recognize other types of visual stimuli and the presence of generally intact intellectual functioning

proximity The principle stating that parts that are close to one another in the visual field are perceived as a whole

psychoanalytic psychology The view that the mind is made up of distinct components, each competing with one another and vying for control of behavior

psychological concept of mind The idea of mind as mental states that cause and explain behavior

psychological inertia A symptom of executive dysfunction characterized by listlessness and an inability to stop some action once it is started

psychologist's fallacy The idea that one person's subjective response to a perception does not guarantee that the same response will exist in the mind of anyone else who experiences the same perception

psychology The scientific study of mind and behavior

psychophysical relations The relation between the psychological mind and the physical body

psychophysics The use of quantitative methods to measure relationships between stimuli, which are physical, and perception, which is psychological

punishment Any consequence that decreases the frequency of a preceding behavior

qualia The felt or experienced character of mental states

quality The characteristic that distinguishes sensations from one another

randomize To randomly allocate experiment participants to the various conditions of the experiment.

random networks Networks where the connections are entirely local and can, therefore, be both short and long distance

random processes Those that exhibit little or no order and, therefore, cannot be predicted. They are described statistically in terms of probabilities instead of certainties.

rationalism The belief in the existence of innate ideas

raw primal sketch An image represented in terms of its distribution of intensity values or areas of light and dark

reagent A substance added to a mixture to produce a particular chemical reaction

reality principle Motivates one to act in a rational and pragmatic fashion

reciprocal altruism A characteristic of early human society involving the sharing of hard-won resources between group members

reconstructive memory When recall of an item from memory is based on guided search and subject to bias by subsequent information

recurrent pathways Those in which information flow can feed back onto itself

reductionism The idea that an understanding of parts can completely account for the behavior of a whole

referent The thing or things in the external world that a representation stands for

reflective The stage in the CogAff architecture model where metacognition comes into play. At this level, one is aware of experiencing an emotion but can choose to act on it or not depending on the circumstances and the use of deliberative processes.

reflex A behavior in which a stimulus triggers sensory neurons that activate intermediary neurons, which in turn activate motor neurons, causing a simple motor response

regulatory model of development A model in which brain mechanisms are only partially determined by genes and operate more slowly, and the parts of the system develop interdependently

rehearsal The repeated activation of a circuit that increases the strength of the connections between its nodes

reinforcement Any event that increases the frequency of a preceding response

representation Something that stands for something else

representativeness heuristic The tendency to judge an item based on its perceived similarity to a category label

repression A defense mechanism that banishes anxiety-arousing thoughts and feelings from consciousness

reticular activating system (RAS) A network of about 100 nuclei that control the brain's overall arousal and alertness level

retrieval The act of accessing needed data from memory and making them available for use

retrieval cues A phenomenon in which an item related to one that was memorized can lead to successful recall

retroactive interference Occurs when information that is learned later interferes with remembering information learned earlier

retrograde amnesia An inability to remember information acquired prior to some traumatic event

reverse engineering The process of starting with an end product and analyzing it to determine its intended function

schema A framework or body of knowledge on some topic; knowledge of how to act as well as the computational processes by which to accomplish the activity (algorithm)

scientific method A process of studying natural phenomena that involves observation and the testing of hypotheses through the use of experiments

script A type of schema that acts as a stored framework or body of knowledge about some topic

seizure An uncontrolled surge of synchronous electrical activity in the brain

selective attention A form of attention that can be focused onto one source of information and away from another

selective exposure The principle by which people seek information not already present that supports a consistency between their attitudes or between their attitudes and behavior

self-disclosure The act of revealing some confidential aspect of ourselves

self-serving bias The tendency to explain our own failures to external causes but to explain other people's failures to internal ones

self-terminating search A memory search process that stops as soon as a positive match between a list item and target occurs

semantic The meaning of a representation

semantic code A memory code based on the inherent meaning of a stimulus

semantic memory A type of declarative memory containing knowledge of facts

semantic network A network where each node has a specific meaning and, therefore, employs local representation of concepts

semantics The rules for deriving meaning in a language

sensitivity to initial conditions A feature of chaotic systems whereby small changes in input can over time produce vastly different outcomes

sensory memory A short-term repository for incoming sensory information

sentence verification A procedure in which participants judge the truth or falsity of sentences by pushing one of two buttons

serial memory search A memory search in which each item in a list is matched against a target, one at a time

serial processors A type of information processing where one computation is performed at a time. The results of one computation typically serve as the input for the next.

sexual division of labor A division of labor between the sexes that is believed to have existed in early human societies where males hunted and females gathered

sexual selection A differential inheritance of traits based on mate selection and competition instead of a change in the environment

shape constancy When a familiar object retains its perceived shape despite radical alterations in how the image is actually being projected onto the retina

shiftability The idea that attention can be shifted back and forth between different sources

similarity The perceptual principle stating parts that are similar in lightness, color, shape, or some other characteristic group together

simple cells Cells in the primary visual cortex that code for oriented line segments

simple ideas Ideas derived through sensory input or simple processes of reflection

single-cell recording A brain recording technique where a fine microelectrode is inserted either into a single neuron or into the extracellular fluid adjacent to it. Changes in that cell's electrical conductivity or its rate of firing are then measured.

small-world network Any network where the number of links between any pair of nodes is few

social cognition The study of how people make sense of other people and of themselves

social dilemmas The types of situations studied by game theorists

social phobia A disorder characterized by fear of public places, social interaction, and being evaluated negatively by others

spandrel An architectural feature formed by the triangular space between the exterior curve of an arch and the rectangular frame enclosing it. In evolution, it is used to designate a by-product of an adaptation that may serve a useful purpose.

sparse distributed coding When a very small group of neurons are connected together to represent a particular object, word, or concept

speech spectrogram A plot that displays the different frequency components of speech and how they change over time

spreading activation The activity that spreads outward from nodes along links to activate other nodes

stability–plasticity dilemma A scenario where a network should be plastic to store novel input patterns. At the same time, it should be stable to protect previously encoded patterns from being erased.

state space A way of visualizing change in chaotic or dynamical systems. Each dimension of the space represents a single variable. A point in this space indicates the current state at any given point in time.

statistical learning The ability for humans and animals to extract statistical regularities from the environment in order to learn. In language acquisition, babies are able to learn very quickly which sounds go together to form words based on the frequency with which those sounds go together.

stereotypes Contain information about people in a specific category. The category can be defined by almost any criteria, such as skin color, sex, profession, or religion.

stigmergy A form of self-organization in which an organism leaves traces of its activity in the environment that then help it (or others) to repeat those activities in the future

stimulus enhancement The explanation that one animal can imitate another simply on the basis of being drawn to the location or behavior without any inference of desired goals

stimulus error Confusing our true experience of an object with a description of the object based on language and past experience

strange attractor A chaotic mathematical system that produces unpredictable values that nonetheless display some regularity in their patterns

stream of consciousness The notion that the mind is a process undergoing continuous flow or change

strong AI View asserts that consciousness can arise from a purely physical process

structured A property of language in which rules specify how symbols can be combined

sublimation A defense mechanism that transforms unacceptable impulses into socially valued motivations

subordinate The most concrete or specific form of conceptual category organization

substance dualism The view that the mind and the body are composed of completely different substances

substantive thought Occurs when the mind slows down, perhaps through the focusing of attention

subsymbolic Processes that operate at a level "underneath" symbols; those mechanisms that give rise to symbolic representations

sulcus A smaller spacing or separation between two areas of brain tissue

sunk-cost fallacy An explanation for the endowment effect by which we prefer to stay with what we have because of the costs we have already invested

superego The aspect of mind that is responsible for our ethical sense of right and wrong; operates on the idealistic principle

superior colliculus A brain area responsible for the moving of visual attention from one object or position in space to another

superordinate The most abstract form of conceptual category organization that encompasses all examples of the concept

Supervised learning algorithm A method for training artificial neural networks in which a "teacher" provides the network with the desired output activation for each input pattern

supervisory attentional system (SAS) A set of schemas that govern only nonroutine actions that require controlled attentional processes. These schemas are designed to be more general purpose and applicable to a wide variety of different problem types.

surface representation A quasi-pictorial representation that occurs in a spatial medium. It depicts an object or scene and underlies our experience of imagery.

sustainability The idea that attention can be maintained over time

symbol A representation that stands for something else

symbol grounding problem An argument against the PSSH that states computer symbols cannot have semantic properties because computer programs are not in bodies that can perceive and act in the world

symbolic dynamics A technique for studying cognitive systems whereby a given region of a state space corresponds to a symbol. The trajectory through different regions corresponds to the processing of one symbol after another.

synaptic cleft The space between neurons across which neurotransmitter molecules travel; typically found between the terminal button of one cell and the dendrite of another

synaptic plasticity A change in the structure or biochemistry of a synapse that occurs during learning

synchrony Coordinated patterns of network activity. The firing rate of distant nodes can synchronize, as can more complicated frequency and timing relationships.

syntax The rules that govern the arrangement of words together in sentences

the teacher A corrective mechanism that compares actual with desired output and generates an error signal

template An internal mental representation of a stimulus to which an image generated from an external stimulus is matched

temporal lobe Located laterally on each hemisphere. It mediates auditory processing, pattern recognition, and language comprehension.

terminal button A bulb-like structure found at the end of the axon. It releases neurotransmitters.

thalamus This brain structure serves as a relay center, forwarding incoming messages from the different senses to parts of the cortex specialized for processing them

theory A set of statements that organizes facts and aids in understanding how the world works

theory of mind (ToM) The ability to understand others' mental states, such as beliefs, desires, and intentions;

to appreciate how these differ from our own; and to use this understanding to guide our behavior when interacting with others

theory of natural selection The theory proposed by Charles Darwin that accounts for changes in animal species over time. It involves species variability, inheritance of traits through reproduction, and selection due to environmental change.

threat–superiority effect The finding that dangerous items are found to be more perceptually salient

threshold The minimum amount of activation required to produce conscious awareness of a stimulus

threshold of excitation The minimal change in a neuron's normal resting electrical state that will initiate the creation of an electrical signal (action potential)

topography A topographical neural representation codes the locations of sensory inputs in a two-dimensional map-like layout; stimuli that are near each other activate neurons that are near each other.

T.O.T. phenomenon This is an acronym for "tip of the tongue," meaning one feels familiarity with an item but cannot quite recall it

trajectory A path through a state space used to measure a system's behavior

transcranial magnetic stimulation (TMS) A method in which a wire coil induces a magnetic field over the brain and can be used to record cortical activity or treat depression

transitive inference The capacity to understand the set of relations between two pairs of items that differ along a continuum (if A > B and B > C, then A > C)

transitive thought The flow of thought that occurs during less focused and more associative forms of reasoning

trial-and-error learning Proceeds from random responses to a behavior that satisfies the requirements of the task

tridimensional theory of feeling A theory stating that all feelings can be characterized by three dimensions: pleasure–displeasure, tension–relaxation, and excitement–depression

two-word stage A stage in language development in which children first produce two-word utterances

typicality effect The phenomenon that human participants are faster to judge stereotypical members as belonging to a category

ultimatum game A laboratory scenario in which people are asked to choose the amount of money they can share with someone else

unconditioned response A response elicited by an unconditioned stimulus

unconditioned stimulus A stimulus that elicits a response on its own

unconscious That aspect of mind of which we are completely unaware

universal grammar The features that are instantiated in the grammars of all natural languages

unsupervised learning algorithm A method for training an artificial neural network that relies on correlations among active neurons and does not need a "teaching signal."

variation Refers to the differences in traits between animals

ventral visual pathway A pathway that travels downward to the temporal lobe and carries data about color and form

verification A stage in problem solving where an insight is confirmed and one checks to see that it produces a correct solution

viewer-centered description An object representation that is particular to the viewer's point of view

visual agnosia A deficit resulting from brain damage in which an individual has difficulty recognizing objects visually

visual buffer The spatial medium of the surface representation. The buffer is a surface matrix consisting of an array of points.

visual code A memory code based on the visual appearance of a stimulus

visual image A mental representation of an object or scene that preserves metric spatial information

visual search The task of identifying a target item located in a field filled with nontarget items, or distracters

visuospatial sketchpad A system in the working memory model that is specialized for the processing of visual information

voluntarism A movement that viewed the mind as consisting of elements and stressed that these elements were assembled into higher level cognitive components through the power of the will

von Economo neurons Neurons in humans and other animal species that allow for the fast transmission of neural signals and that may be an indicator for intelligence

Wason selection task A task designed to measure a person's logical thinking ability. It involves applying the abstract rules of logic to a specific example.

weak AI Consciousness is itself either not a physical process, and so can never be reproduced, or is a physical process but such a complex one that we will never be able to duplicate it artificially

weights Values on links that determine the strength of the connection between nodes. They can range in value from –1.0 to +1.0.

Wernicke's aphasia A language deficit in which speech production is intact but there are impairments in comprehension

Wernicke's area A region in the left temporal lobe that underlies language comprehension. Damage to this area results in Wernicke's aphasia.

whole-report condition A condition in iconic memory studies in which the task is to remember the entire letter display

Williams syndrome A genetic disease caused by a deletion on chromosome 7. These individuals manifest hypersociability. They show an exaggerated interest in other people and increased expressiveness and social communicability.

working memory A short-term memory store used to represent and transform information

REFERENCES

Abell, F., Krams, M., Ashburner, J., Passingham, R., Friston, K., Frackowiak, R., . . . Frith, U. (1999). The neuroanatomy of autism: A voxel-based whole brain analysis of structural scans. *NeuroReport*, *10*(8), 1647–1651.

Able, K. P. (1996). The debate over olfactory navigation by homing pigeons. *Journal of Experimental Biology*, *199*, 121–124.

Aboud, F. (2003). The formation of in-group favoritism and out-group prejudice in young children: Are they distinct attitudes? *Developmental Psychology*, *39*(1), 48–60.

Abrahams, B. S., & Geschwind, D. H. (2008). Advances in autism genetics: On the threshold of a new neurobiology. *Nature Reviews Genetics*, *9*(5), 341–355.

Ackley, D., & Littman, M. (1992). Interactions between learning and evolution. In C. Langton (Ed.), *Artificial life II* (pp. 487–509). Redwood City, CA: Addison-Wesley.

Adamatzky, A. (2018). Towards fungal computer. *Interface Focus*, *8*(6), 20180029.

Adams, F., & Aizawa, K. (2009). Why the mind is still in the head. In M. Aydede & P. Robbins (Eds.), *The Cambridge handbook of situated cognition* (pp. 78–95). Cambridge, England: Cambridge University Press.

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, *4*(3), 165–178.

Adolphs, R., & Tranel, D. (1999). Preferences for visual stimuli following amygdala damage. *Journal of Cognitive Neuroscience*, *11*(6), 610–616.

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. *Nature*, *393*, 470–474.

Alais, D., & Blake, R. (2005). *Binocular rivalry*. Cambridge: MIT Press.

Allman, J. M., Hakeem, A., Erwin, J. M., Nimchinsky, E., & Hof, P. (2001). The evolution of an interface between emotion and cognition. *Annals of the New York Academy of Sciences*, *935*, 107–117.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439.

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238.

Ambady, N., Koo, J., Rosenthal, R., & Winograd, C. (2002). Physical therapists' nonverbal communication predicts geriatric patients' health outcomes. *Psychology and Aging*, *17*(3), 443–452.

Anderson, J. R. (1980). *Cognitive psychology and its implications* (2nd ed.). San Francisco, CA: W. H. Freeman.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science*, 2, 396–408.

Anderson, M. L. (2014). *After phrenology: Neural reuse and the interactive brain*. Cambridge: MIT Press.

Anderson, M. L., & Finlay, B. L. (2014). Allocating structure to function: The strong links between neuroplasticity and natural selection. *Frontiers in Human Neuroscience*, 7, 918.

Anderson, S. E., Huette, S., Matlock, T., & Spivey, M. J. (2010). On the temporal dynamics of negated perceptual simulations. In M. Turner, F. Parrill, & V. Tobin (Eds.), *Meaning, form, and body* (pp. 1–20). Stanford, CA: CSLI Publications.

Andrews, P. W., Thomson, J., & Anderson, J. (2009). The bright side of being blue: Depression as an adaptation for analyzing complex problems. *Psychological Review*, *116*(3), 620–654.

Angell, J. R. (1907). The province of functional psychology. *Psychological Review*, *14*, 61–91.

Arndt, T. L., Stodgell, C. J., & Rodier, P. M. (2005). The teratology of autism. *International Journal of Developmental Neuroscience*, *23*(2–3), 189–199.

Ashcraft, M. H. (2002). *Cognition* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Astafiev, S., Shulman, G., Stanley, C., Snyder, A., Essen, D., & Corbetta, M. (2003). Functional organization of human intraparietal and frontal cortex for attending, looking, and pointing. *Journal of Neuroscience*, *23*, 4689–4699.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American*, *225*, 82–90.

Au, K., Chan, F., Wang, D., & Vertinsky, I. (2003). Mood in foreign exchange trading: Cognitive processes and performance. *Organizational Behavior and Human Decision Processes*, *91*, 322–328.

Au, T. K. (1983). Chinese and English counterfactuals: The Sapir-Whorf hypothesis revisited. *Cognition*, *15*(1–3), 155–187.

Au, T. K. (1984). Counterfactuals: In reply to Alfred Bloom. *Cognition*, *17*(3), 289–302.

Averbach, E., & Sperling, G. (1961). Short-term storage and information in vision. In C. Cherry (Ed.), *Information theory* (pp. 196–211). London, England: Butterworth.

Axelrod, R. (1980). Effective choice in the prisoner's dilemma. *Journal of Conflict Resolution*, *24*(1), 3–25.

Axmacher, N., Mormann, F., Fernandez, G., Elger, C. E., & Fell, J. (2006). Memory formation by neural synchronization. *Brain Research Reviews*, *52*, 170–182.

Babad, E., Bernieri, F., & Rosenthal, R. (1989). When less information is more informative: Diagnosing teacher expectations from brief samples of behaviour. *British Journal of Educational Psychology*, *59*(3), 281–295.

Bach-y-Rita, P., Collins, C. C., Saunders, F., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, *221*, 963–964.

Bachevalier, J. (1994). Medial temporal lobe structures and autism: A review of clinical and experimental findings. *Neuropsychologia*, *32*(6), 627–648.

Bachevalier, J., Málková, L., & Mishkin, M. (2001). Effects of selective neonatal temporal lobe lesions on socioemotional behavior in infant rhesus monkeys (*Macaca mulatta*). *Behavioral Neuroscience*, *115*(3), 545–559.

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556–559.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, *113*, 1–29.

Baig, M. M., Gholamhosseini, H., & Harrison, M. J. (2012). *Fuzzy logic based smart anesthesia monitoring system in the operation theatre*. Retrieved from http://www.wseas.org/multimedia/journals/circuits/2012/53-598.pdf

Bailey, A., Phillips, W., & Rutter, M. (1996). Autism: Towards an integration of clinical, genetic, neuropsychological, and neurobiological perspectives. *Journal of Child Psychology and Psychiatry*, *37*(1), 89–126.

Baker, M. C. (2001). *The atoms of language*. New York, NY: Basic Books.

Balasubramaniam R. (2013). On the control of unstable objects: the dynamics of human stick balancing. In M. Richardson, M. Riley, & K. Shockley (Eds.), *Progress in motor control* (pp. 149–168). Berlin, Germany: Springer.

Baldwin, D. A. (1995). Understanding the link between joint attention and language. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 131–158). Hillsdale, NJ: Erlbaum.

Banich, M. T. (2004). *Cognitive neuroscience and neuropsychology*. Boston, MA: Houghton Mifflin.

Barabási, A. L. (2009). Scale-free networks: A decade and beyond. *Science*, *325*(5939), 412–413.

Baran, P. (1964). *Introduction to distributed communications networks* (Report Memorandum RM-3240-PR). Santa Monica, CA: RAND.

Barinaga, M. (1992). How scary things get that way. *Science*, *258*, 887–888.

Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences*, *6*(6), 248–254.

Baron-Cohen, S., Cox, A., Baird, G., Swettenham, J., & Nightingale, N. (1996). Psychological markers in the detection of autism in infancy in a large population. *British Journal of Psychiatry*, *168*(2), 158–163.

Baron, J. (2000). *Thinking and deciding* (3rd ed.). New York, NY: Cambridge University Press.

Barrett, L. F. (2006). Are emotions natural kinds? *Perspectives on Psychological Science*, *1*(1), 28–58.

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Boston, MA: Houghton Mifflin Harcourt.

Barsalou, L. W. (1983). Ad hoc categories. *Memory & Cognition*, *11*(3), 211–227.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660.

Barsalou, L. W. (2016). On staying grounded and avoiding quixotic dead ends. *Psychonomic Bulletin & Review*, *23*(4), 1122–1142.

Bauman, M., & Kemper, T. (2007). The neuroanatomy of the brain in autism: Current thoughts and future directions. In J. M. Pérez, P. M. González, M. L. Comí, & C. Nieto (Eds.), *New developments in autism: The future is today* (pp. 259–267). London, England: Jessica Kingsley.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, *117*, 497–529.

Bavelier, D., Corina, D., Jezzard, P., Clark, V., Karni, A., Lalwani, A., . . . Neville, H. J. (1998). Hemispheric specialization for English and ASL: Left invariance–right variability. *NeuroReport*, *9*(7), 1537–1542.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, (53), 370–418.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*, 1293–1294.

Bechtel, W., & Abrahamsen, A. (1991). *Connectionism and the mind: An introduction to parallel processing in networks*. Cambridge, MA: Basil Blackwell.

Bechtel, W., & Abrahamsen, A. (2006). Phenomena and mechanisms: Putting the symbolic, connectionist, and dynamical systems debate in broader perspective. In R. Stainton (Ed.), *Contemporary debates in cognitive science* (pp. 159–186). Oxford, England: Basil Blackwell.

Becker, M. W., Pashler, H., & Anstis, S. M. (2000). The role of iconic memory in change detection tasks. *Perception*, *29*(3), 273–286.

Beer, R. D. (1990). *Intelligence as adaptive behavior: An experiment in computational neuroethology*. New York, NY: Academic Press.

Beer, R. D., & Williams, P. L. (2015). Information processing and dynamics in minimally cognitive agents. *Cognitive Science*, *39*(1), 1–38.

Behrmann, M., & Geng, J. J. (2002). What is 'left' when all is said and done? Spatial coding and hemispatial neglect. In H.-O. Karnath, D. Milner, & G. Vallar (Eds.), *The cognitive and neural bases of spatial neglect* (pp. 85–100). Oxford, England: Oxford University Press.

Behrmann, M., & Plaut, D. C. (2013). Distributed circuits, not circumscribed centers, mediate visual recognition. *Trends in Cognitive Sciences*, *17*(5), 210–219.

Behrmann, M., Geng, J. J., & Shomstein, S. (2004). Parietal cortex and attention. *Current Opinion in Neurobiology*, *14*(2), 212–217.

Bellugi, U., Lichtenberger, L., Jones, W., Lai, Z., & St. George, M. (2000). The neurocognitive profile of Williams syndrome: A complex pattern of strengths and weaknesses. *Journal of Cognitive Neuroscience*, *12*, 7–29.

Bellugi, U., Wang, P. P., & Jernigan, T. L. (1994). Williams syndrome: An unusual neuropsychological profile. In S. H. Broman & J. Grafman (Eds.), *Atypical cognitive deficits in developmental disorders: Implications for brain function* (pp. 23–56). Hillsdale, NJ: Erlbaum.

Bennett, W., Pickard, V., Iozzi, D., Schroeder, C., Lagos, T., & Caswell, C. (2004). Managing the public sphere: Journalistic construction of the great globalization debate. *Journal of Communication*, *54*(3), 437–455.

Bergen, B. K. (2004). The psychological reality of phonaesthemes. *Language*, *80*(2), 290–311.

Bergen, B. K., Lindsay, S., Matlock, T., & Narayanan, S. (2007). Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science*, *31*(5), 733–764.

Berlin, B., & Kay, P. (1969). *Basic color terms: Their universality and evolution*. Berkeley: University of California Press.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115–147.

Biederman, I. (1995). Visual object recognition. In M. Kosslyn & D. N. Osherson (Eds.), *An invitation to cognitive science: Visual cognition* (2nd ed., Vol. 2, pp. 121–165). Cambridge: MIT Press.

Binder, J. R., Westbury, C. F., McKiernan, K. A., Possing, E. T., & Medler, D. A. (2005). Distinct brain systems for processing concrete and abstract concepts. *Journal of Cognitive Neuroscience*, *17*(6), 905–917.

Birbaumer, N., Grodd, W., Diedrich, O., Klose, U., Erb, M., Lotze, M., . . . Flor, H. (1998). fMRI reveals amygdala activation to human faces in social phobics. *NeuroReport*, *9*(6), 1223–1226.

Bishop, D. V. M. (1997). *Uncommon understanding: Development and disorders of language comprehension in children*. Hove, England: Psychology Press.

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, *2*, 561–567.

Blakemore, S. J., Winston, J., & Frith, U. (2004). Social cognitive neuroscience: Where are we heading? *Trends in Cognitive Sciences*, *8*, 216–222.

Blanchette, I., & Dunbar, K. (2001). Analogy use in naturalistic settings: The influence of audience, emotion, and goals. *Memory & Cognition*, *29*(5), 2001.

Bloom, A. H. (1981). *The linguistic shaping of thought: A study in the impact of language on thinking in China and the West*. Mahwah, NJ: Erlbaum.

Bloom, P. (Ed.). (1994). *Language acquisition*. Cambridge: MIT Press.

Bluff, L. A., Weir, A. A. S., & Rutz, J. H. (2007). Tool-related cognition in New Caledonian crows. *Comparative Cognition & Behavior Reviews*, *2*, 1–25.

Boden, M. (1990). Escaping from the Chinese room. In M. Boden (Ed.), *The philosophy of artificial intelligence* (pp. 88–104). Oxford, England: Oxford University Press.

Bolhuis, J. J., & van Kampen, H. S. (1988). Serial position curves in spatial memory of rats: Primacy and recency effects. *Quarterly Journal of Experimental Psychology B*, *40*, 135–149.

Borghi, A. M., Barca, L., Binkofski, F., Castelfranchi, C., Pezzulo, G., & Tummolini, L. (2018). Words as social tools: Language, sociality and inner grounding in abstract concepts. *Physics of Life Reviews*, *29*, 120–153.

Borghi, A. M., Flumini, A., Cimatti, F., Marocco, D., & Scorolli, C. (2011). Manipulating objects and telling words: A study on concrete and abstract words acquisition. *Frontiers in Psychology*, *2*, 15.

Boroditsky, L. (2001). Does language shape thought? Mandarin and English speakers' conceptions of time. *Cognitive Psychology*, *43*(1), 1–22.

Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 61–79). Cambridge: MIT Press.

Bortfeld, H., Fava, E., & Boas, D. A. (2009). Identifying cortical lateralization of speech processing in infants using near-infrared spectroscopy. *Developmental Neuropsychology*, *34*(1), 52–65.

Boster, J., & D'Andrade, R. (1989). Natural and human sources of cross-cultural agreement in ornithological classification. *American Anthropologist*, *91*, 132–142.

Bostrom, N. (2016). *Superintelligence: Paths, dangers, strategies*. Oxford, England: Oxford University Press.

Bower, G. H. (1981). Mood and memory. *American Psychologist*, *36*, 129–148.

Bower, G. H., Monteiro, K. P., & Gilligan, S. G. (1978). Emotional mood as a context of learning and recall. *Journal of Verbal Learning and Verbal Behavior*, *17*, 573–585.

Bowlby, J. (1967). *Attachment and loss: Attachment* (Vol. 1). New York, NY: Basic Books.

Boyd, R., & Richerson, P. (1996). Why culture is common, but cultural evolution is rare. *Proceedings of the British Academy*, *88*, 77–93.

Braitenberg, V. (1986). *Vehicles: Experiments in synthetic psychology*. Cambridge: MIT Press.

Bravais, A. (1949). *On the systems formed by points regularly distributed on a plane or in space*. Buffalo, NY: Crystallographic Society of America.

Bray, D. (2009). *Wetware: A computer in every living cell*. New Haven, CT: Yale University Press.

Breazeal, C. (2002). *Designing sociable robots*. Cambridge: MIT Press.

Breazeal, C., & Brooks, R. (2005). Robot emotions: A functional perspective. In J. M. Fellows & M. A. Arbib (Eds.), *Who needs emotions? The brain meets the robot* (pp. 271–310). Oxford, England: Oxford University Press.

Brewer, M. B., & Brown, R. J. (1998). Intergroup relations. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 2, 4th ed., pp. 554–594). New York, NY: McGraw-Hill.

Brewer, M. M., & Harasty Feinstein, A. S. (1999). Dual processes in the cognitive representation of persons and social categories. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 225–270). New York, NY: Guilford Press.

Broadbent, D. E. (1958). *Perception and communication*. London, England: Pergamon Press.

Brock, T. C., & Balloun, J. L. (1967). Behavioral receptivity to dissonant information. *Journal of Personality and Social Psychology*, *6*, 413–428.

Brock, T. C., Albert, S. M., & Becker, L. A. (1970). Familiarity, utility, and supportiveness as determinants of information receptivity. *Journal of Personality and Social Psychology*, *14*, 292–301.

Brockman, J. (2019). *Possible minds: Twenty-five ways of looking at AI*. New York, NY: Penguin.

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139–159.

Brooks, R. A. (1999). *Cambrian intelligence: The early history of the new AI*. Cambridge: MIT Press.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology: Cognitive development* (Vol. 3, 4th ed., pp. 77–166). New York, NY: Wiley.

Brown, A. S. (1991). A review of the tip-of-the-tongue experience. *Psychological Bulletin*, *109*(2), 204–223.

Brown, J. A. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, *10*, 12–21.

Brown, J. D., & Rogers, R. J. (1991). Self-serving attributions: The role of physiological arousal. *Personality and Social Psychology Bulletin*, *17*(5), 501–506.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review*, *18*(6), 1189–1196.

Brown, R. J. (2000). Social identity theory: Past achievements, current problems, and future challenges. *European Journal of Social Psychology*, *30*, 745–778.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition*, *5*, 73–93.

Bruce, C., Desimone, R., & Gross, C. G. (1981). Visual properties of neurons in a polysensory area in the superior temporal sulcus of the macaque. *Journal of Neurophysiology*, *46*, 369–384.

Brugman, C., & Lakoff, G. (1988). Cognitive topology and lexical networks. In S. Small, G. Cottrell, & M. Tanenhaus (Eds.), *Lexical ambiguity resolution* (pp. 477–508). San Mateo, CA: Morgan Kaufmann.

Brunet, E., Sarfati, Y., Hardy-Bayle, M. C., & Decety, J. (2000). A PET investigation of the attribution of intentions with a nonverbal task. *NeuroImage*, *11*, 157–166.

Buchanan, M. (2002). *Nexus: Small worlds and the groundbreaking theory of networks*. New York, NY: W. W. Norton.

Burton, A. M., Young, A. W., Bruce, V., Johnston, R. A., & Ellis, A. W. (1991). Understanding covert recognition. *Cognition*, *39*, 129–166.

Buss, D. M. (1999). *Evolutionary psychology: The new science of the mind*. Boston, MA: Allyn & Bacon.

Byrne, R. W. (1995). *The thinking ape*. Oxford, England: Clarendon Press.

Cacioppo, J., Crites, S., & Gardner, W. (1996). Attitudes to the right: Evaluative processing is associated with lateralized late positive event-related brain potentials. *Personality and Social Psychology Bulletin*, *22*(12), 1205–1219.

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or doing? Influence of visual and motor familiarity in action observation. *Current Biology*, *16*(19), 1905–1910.

Camerer, C. F. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton, NJ: Princeton University Press.

Cannon, W. B. (1927). The James-Lange theory of emotion: A critical examination and an alternative theory. *American Journal of Psychology*, *39*, 10–124.

Caporale, N., & Dan, Y. (2008). Spike timing–dependent plasticity: A Hebbian learning rule. *Annual Review Neuroscience*, *31*, 25–46.

Carabotti, M., Scirocco, A., Maselli, M. A., & Severi, C. (2015). The gut-brain axis: Interactions between enteric microbiota, central and enteric nervous systems. *Annals of Gastroenterology*, *28*(2), 203–209.

Carey, B. (2012, May 16). Paralyzed, moving a robot with their minds. *New York Times*. Retrieved from http://www.nytimes.com/2012/05/17/science/bodies-inert-they-moved-a-robot-with-their-minds.html

Carlson, S., & Moses, L. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*(4), 1032–1053.

Carr, H. A. (1925). *Psychology*. New York, NY: Longmans, Green.

Carroll, J. B. (1956). *Language, thought, and reality: Selected writings of Benjamin Lee Whorf*. Cambridge: MIT Press.

Carter, C., & Rice, C. (1997). Acquisition and manifestation of prejudice in children. *Journal of Multicultural Counseling and Development*, *25*(3), 185–194.

Casasanto, D., & Lupyan, G. (2015). All concepts are ad hoc concepts. In E. Margolis & S. Laurence (Eds.), *The conceptual mind: New directions in the study of concepts* (pp. 543–566). Cambridge: MIT Press.

Castano, E., & Yzerbyt, V. Y. (1998). The highs and lows of group homogeneity. *Behavioral Processes*, *42*, 219–238.

Cawsey, A. (1998). *The essence of artificial intelligence*. Harlow, England: Prentice Hall.

Chakraborty, R. C. (2008, March 16). Artificial intelligence technologies. Retrieved from myreaders.wordpress.com

Chalmers, D. (1996). *The conscious mind*. Oxford, England: Oxford University Press.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(3), 687–696.

Chapanis, A. (1965). Color names for color space. *American Scientist*, *53*, 327–346.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). New York, NY: Academic Press.

Chen, H., Tan, H., Kuntz, A., Bansal, M., & Alterovitz, R. (2020, May). Enabling robots to understand incomplete natural language instructions using commonsense reasoning. In 2020 IEEE International Conference on Robotics and Automation (ICRA) (pp. 1963–1969). IEEE.

Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world*. Chicago, IL: University of Chicago Press.

Cherkassky, V., Kana, R., Keller, T., & Just, M. (2006). Functional connectivity in baseline resting state network in autism. *NeuroReport*, *17*, 1687–1690.

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, *25*, 975–979.

Chomsky, N. (1957). *Syntactic structures*. The Hague, Netherlands: Mouton.

Chomsky, N. (1959). A review of BF Skinner's Verbal Behavior. *Language*, *35*(1), 26–58.

Chomsky, N. (1980). *Rules and representations*. New York, NY: Columbia University Press.

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York, NY: Praeger.

Choset, H., Lynch, K. M., Hutchinson, S., Kantor, G., Burgard, W., Kavraki, L. E., & Thrun, S. (2005). *Principles of robot motion: Theory, algorithms, and implementations*. Cambridge: MIT Press.

Christianson, S.-A., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Emotion and Cognition*, *5*, 81–108.

Chrysikou, E. G., Hamilton, R. H., Coslett, H. B., Datta, A., Bikson, M., & Thompson-Schill, S. L. (2013). Noninvasive transcranial direct current stimulation over the left prefrontal cortex facilitates cognitive flexibility in tool use. *Cognitive Neuroscience*, *4*(2), 81–89.

Church, A. (1936). An unsolvable problem of elementary number theory. *American Journal of Mathematics*, *58*, 345–363.

Churchland, P. M. (1981). Eliminativism and the propositional attitudes. *Journal of Philosophy*, *78*, 67–90.

Churchland, P. M. (1995). *The engine of reason, the seat of the soul: A philosophical journey into the brain*. Cambridge: MIT Press.

Churchland, P. S., Koch, C., & Sejnowski, T. J. (1990). What is computational neuroscience? In E. L. Schwartz (Ed.), *Computational neurosciences* (pp. 38–45). Cambridge: MIT Press.

Cialdini, R. (1993). *Influence: Science and practice* (3rd ed.). New York, NY: HarperCollins.

Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science*. New York, NY: Oxford University Press.

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. New York, NY: Oxford University Press.

Clark, A. (2015). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford, England: Oxford University Press.

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis, 58*, 7–19.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York, NY: Harcourt Brace Jovanovich.

Clark, V. P., Keil, K., Maisog, J. M., Courtney, S. M., Ungerleider, L. G., & Haxby, J. V. (1996). Functional magnetic resonance imaging of human visual cortex during face matching: A comparison with positron emission tomography. *NeuroImage*, *4*, 1–15.

Clore, G. L., Gasper, K., & Garvin, E. (2001). Affect as information. In J. P. Forgas (Ed.), *Handbook of affect and social cognition* (pp. 121–144). Mahwah, NJ: Erlbaum.

Cohen, D. (1979). *J. B. Watson: The founder of behaviourism: A biography*. London, England: Routledge & Kegan Paul.

Colby, C. L., Duhamel, J.-R., & Goldberg, M. E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of Neurophysiology*, *76*, 2841–2852.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240–247.

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*, *116*(3), 457–475.

Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language*, *26*(5), 527–538.

Conrad, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, *55*, 75–84.

Cook, V. J. (1988). *Chomsky's universal grammar: An introduction*. Cambridge, MA: Basil Blackwell.

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 75–176). New York, NY: Academic Press.

Corkin, S. (1984). Lasting consequences of bilateral medial temporal lobectomy: Clinical course and experimental findings in H. M. *Seminars in Neurology*, *4*, 249–259.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (pp. 163–228). New York, NY: Oxford University Press.

Coull, J. T., & Nobre, A. C. (1998). Where and when to pay attention: The neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *Journal of Neuroscience*, *18*, 7426–7435.

Cowen, A. S., & Keltner, D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *Proceedings of the National Academy of Sciences*, *114*(38), E7900–E7909.

Craik, K. (1943). *The nature of explanation*. Cambridge, England: Cambridge University Press.

Crick, F., & Koch, C. (1995). Are we aware of neural activity in primary visual cortex? *Nature*, *375*, 121–123.

Crystal, D. (1987). *The Cambridge encyclopedia of language*. Cambridge, MA: Cambridge University Press.

Cuddy, A. J. C., Fiske, S. R., & Glick, P. (2007). The BIAS map: Behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, *92*, 631–648.

Cummins, D. D. (1999, June). Early emergence of cheater detection in human development. Paper presented at the annual meeting of the Human Behavior and Evolution Society, Salt Lake City, UT.

Cusack, R., Carlyon, R. P., & Robertson, I. H. (2000). Neglect between but not within auditory objects. *Journal of Cognitive Neuroscience*, *12*(6), 1056–1065.

Dale, R., Kehoe, C., & Spivey, M. J. (2007). Graded motor responses in the time course of categorizing atypical exemplars. *Memory & Cognition*, *35*(1), 15–28.

Damasio, A. (1994). *Descartes' error: Emotion, reason, and the human brain*. New York, NY: Putnam Books.

Damasio, A. R. (1999). *The feeling of what happens: Body and emotion in the making of consciousness*. New York, NY: Harcourt Brace.

Darwin, C. (1859). *On the origin of species by means of natural selection*. London, England: John Murray.

Darwin, C. (1998). *The expression of emotion in man and animals* (P. Ekman, Ed.; 3rd ed.). New York, NY: Oxford University Press. (Original work published 1872)

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, *3*, 255–267.

Davidson, R. J., Scherer, K. R., & Goldsmith, H. H. (2003). *Handbook of affective sciences*. New York, NY: Oxford University Press.

Davies, I. R. L., & Corbett, G. G. (1997). A cross-cultural study of colour grouping: Evidence for weak linguistic relativity. *British Journal of Psychology*, *88*(3), 493–517.

Davies, I. R. L., Sowden, P. T., Jerrett, D. T., Jerrett, T., & Corbett, G. G. (1998). A cross-cultural study of English and Setswana speakers on a colour triads task: A test of the Sapir-Whorf hypothesis. *British Journal of Psychology*, *89*(1), 1–15.

Davis, H. (1992). Discrimination of the number three by a raccoon (*Procyon lotor*). *Animal Learning and Behavior*, *12*, 409–413.

Davis, S., & Ludvigson, H. (1995). Additional data on academic dishonesty and a proposal for remediation. *Teaching of Psychology*, *22*(2), 119–121.

Davis, S., Grover, C., Becker, A., & McGregor, L. (1992). Academic dishonesty: Prevalence, determinants, techniques, and punishments. *Teaching of Psychology*, *19*(1), 16–20.

Dawson, M. R. W. (1998). *Understanding cognitive science*. Malden, MA: Blackwell.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R. J. (2006). Frames, biases, and rational decision making in the human brain. *Science*, *313*, 684–687.

de Villiers, J., & de Villiers, P. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 191–228). Hove, England: Psychology Press.

De Waal, F. (2016). *Are we smart enough to know how smart animals are?* New York, NY: W. W. Norton.

De Waal, F. B. M. (1997a). Are we in anthropodenial? *Discover*, *18*(7), 50–53.

De Waal, F. B. M. (1997b). Food-transfers through mesh in brown capuchins. *Journal of Comparative Psychology*, *111*, 370–378.

Decety, J., & Jackson, P. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, *3*(2), 406–412.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage*, *24*, 21–33.

Deisseroth, K. (2010). Controlling the brain with light. *Scientific American*, *303*, 48–55.

DeLancey, C. (2002). *Passionate engines: What emotions reveal about the mind and artificial intelligence*. Oxford, England: Oxford University Press.

Dennett, D. (1991). *Consciousness explained*. Boston, MA: Little, Brown.

Dennett, D. (1998). The practical requirements for making a conscious robot. In *Brainchildren: Essays on designing minds* (pp. 154–170). Cambridge: MIT Press.

Deregowski, J. B. (1989). Real space and represented space: Cross-cultural perspectives. *Behavioral and Brain Sciences*, *12*, 51–119.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Desimone, R., Albright, T. D., Gross, C. D., & Bruce, C. (1984). Stimulus-selective responses of inferior temporal neurons in the macaque. *Journal of Neuroscience*, *4*, 2051–2062.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*, 80–90.

Dijkstra, K., Kaschak, M. P., & Zwaan, R. A. (2007). Body posture facilitates retrieval of autobiographical memories. *Cognition*, *102*(1), 139–149.

Dinstein, I., Gardner, J., Jazayeri, M., & Heeger, D. (2008). Executed and observed movements have different distributed representations in human aIPS. *Journal of Neuroscience*, *28*(44), 11231–11239.

Dion, K., & Stein, S. (1978). Physical attractiveness and interpersonal influence. *Journal of Experimental Social Psychology*, *14*(1), 97–108.

Dolan, R. (2002). Emotion, cognition, and behavior. *Science*, *298*, 1191–1194.

Dotov, D. G., Nie, L., & Chemero, A. (2010). A demonstration of the transition from ready-to-hand to unready-to-hand. *PLoS One*, *5*(3), e9433.

Doupe, A. J., & Kuhl, P. K. (1999). Birdsong and human speech: Common themes and mechanisms. *Annual Review of Neuroscience*, *22*, 567–631.

Dover, G. (2000). Anti-Dawkins. In H. Rose & S. Rose (Eds.), *Arguments against evolutionary psychology* (pp. 55–77). New York, NY: Harmony Books.

Dreyfus, H. (1965). *Alchemy and AI*. Santa Monica, CA: RAND.

Dreyfus, H. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. Oxford, England: Blackwell.

Dreyfus, H. L. (1992). *What computers still can't do* (3rd ed.). Cambridge: MIT Press.

Driver, J., & Mattingley, J. B. (1998). Parietal neglect and visual awareness. *Nature Neuroscience Review*, *1*(1), 17–22.

Dronkers, N. F., Redfern, B. B., & Ludy, C. A. (1995). Lesion localization in chronic Wernicke's aphasia. *Brain and Language*, *51*(1), 62–65.

Dronkers, N. F., Shapiro, J. K., Redfern, B., & Knight, R. T. (1992). The role of Broca's area in Broca's aphasia. *Journal of Clinical and Experimental Neuropsychology*, *14*, 52–53.

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification*. New York, NY: Wiley-Interscience.

Dunbar, R. (1996). *Grooming, gossip, and the evolution of language*. Cambridge, MA: Harvard University Press.

Duncan, B. (1976). Differential social perception and attribution of intergroup violence: Testing the lower limits of stereotyping of Blacks. *Journal of Personality and Social Psychology*, *34*(4), 590–598.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458.

Duncker, K. (1945). On problem solving. *Psychological Monographs*, *58*(5, Whole No. 270).

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature Neuroscience*, *3*(11s), 1184–1191.

Duval, F. (2017). *Deep learning for beginners: Practical guide with python and Tensorflow*. CreateSpace Publishing, Amazon Kindle Direct Publishing.

Dworkin, B. R. (1993). *Learning and physiological regulation*. Chicago, IL: University of Chicago Press.

Eastwood, J. D., Smilek, D., & Merikle, P. M. (2001). Differential attentional guidance by unattended faces expressing positive and negative emotion. *Perception and Psychophysics*, *63*, 1004–1013.

Edelman S. (1999). *Representation and recognition in vision*. Cambridge, MA: MIT Press.

Edmiston, P., & Lupyan, G. (2017). Visual interference disrupts visual knowledge. *Journal of Memory and Language*, *92*, 281–292.

Edwards, F. A. (1995). LTP: A structural model to explain the inconsistencies. *Trends in Neuroscience*, *18*, 250–255.

Ekman, P. (1989). The argument and evidence about universals in facial expressions of emotion. In H. Wagner & A. Manstead (Eds.), *Handbook of social psychophysiology* (pp. 143–164). Chichester, England: Wiley.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*, 124–129.

Elia, J., Ambrosini, P. J., & Rapoport, J. L. (1999). Treatment of attention-deficit hyperactivity disorder. *New England Journal of Medicine*, *340*, 780–788.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, *7*(2), 195–225.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*(4), 547–582.

Elman, J. L., Bates, L., Johnson, M., Karmiloff-Smith, A., Parisi, D. P. K., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge: MIT Press.

Ember, C. R., & Ember, M. (1985). *Cultural anthropology* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

Engel, A. K., König, P., Kreiter, A. K., Schillen, T. B., & Singer, W. (1992). Temporal coding in the visual cortex: New vistas on integration in the nervous system. *Trends in Neurosciences*, *15*, 218–226.

Erickson, B., Lind, E., Johnson, B., & O'Barr, W. (1978). Speech style and impression formation in a court setting: The effects of "powerful" and "powerless" speech. *Journal of Experimental Social Psychology*, *14*(3), 266–279.

Everett, D. L. (2012). *Language: The cultural tool*. New York, NY: Vintage.

Fajardo, C., Escobar, M. I., Buritica, E., Arteaga, G., Umbarila, J., Casanova, M. F., & Pimienta, H. (2008). Von Economo neurons are present in the dorsolateral prefrontal cortex of humans. *Neuroscience Letters*, *435*(3), 215–218.

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review*, *95*, 307–317.

Farah, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. Cambridge: MIT Press.

Faraone, S. V., Doyle, A. E., Mick, E., & Biederman, J. (2001). Meta-analysis of the association between the 7-repeat allele of the dopamine D4 receptor gene and attention-deficit hyperactivity disorder. *American Journal of Psychiatry*, *158*, 1052–1057.

Farmer, T. A., Christiansen, M. H., & Monaghan, P. (2006). Phonological typicality influences on-line sentence comprehension. *Proceedings of the National Academy of Sciences*, *103*(32), 12203–12208.

Favela, L. H. (2020). Cognitive science as complexity science. *Wiley Interdisciplinary Reviews: Cognitive Science*, 11(4), e1525.

Favela, L. H., & Chemero, A. (2015). The animal-environment system. In Y. Coelllo & M. H. Fischer (Eds.), *Foundations of embodied cognition* (Vol. 1, pp. 59–74). New York, NY: Routledge.

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–110). New York, NY: Academic Press.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition and Emotion*, *15*(2), 115–141.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*(6), 1013–1027.

Feingold, A. (1991). Sex differences in the effects of similarity and physical attractiveness on opposite-sex attraction. *Basic and Applied Social Psychology*, *12*(3), 357–367.

Feltz, A., & Cokely, E. T. (2009). Do judgments about freedom and responsibility depend on who you are? Personality differences in intuitions about compatibilism and incompatibilism. *Consciousness and Cognition*, *18*, 342–350.

Fernald, A. (1989). Intonation and communicative intent in mother's speech to infants: Is the melody the message? *Child Development*, *60*, 1497–1510.

Fiedler, K. (2001). Affective states trigger processes of assimilation and accommodation. In L. L. Martin & G. L. Clore (Eds.), *Theories of mood and cognition: A user's guidebook* (pp. 85–98). Mahwah, NJ: Erlbaum.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field." *Vision Research*, *33*(2), 173–193.

Field, M. J. (Ed.). (1996). *Telemedicine: A guide to assessing telecommunications for health care*. Washington, DC: National Academies Press.

Fink, G. R., Dolan, R. J., Halligan, P. W., Marshall, J. C., & Firth, C. D. (1997). Space-based and object-based visual attention: Shared and specific neural domains. *Brain*, *120*, 2013–2028.

Finke, R. A., & Kosslyn, S. M. (1980). Mental imagery acuity in the peripheral visual field. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 126–139.

Finlay, B. L., & Workman, A. D. (2013). Human exceptionalism. *Trends in Cognitive Sciences*, *17*(5), 199–201.

Fiske, S. T., & Taylor, S. E. (2008). *Social cognition: From brains to culture*. New York, NY: McGraw-Hill.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively followed from preceded status and competition. *Journal of Personality and Social Psychology*, *82*, 878–902.

Fiske, S. T., Lin, M. H., & Neuberg, S. L. (1999). The continuum model: Ten years later. In S. Chaiken & Y. Trope (Eds.), *Dual process theories in social psychology* (pp. 231–254). New York, NY: Guilford Press.

Fletcher, P. C., Happé, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: A functional imaging study of "theory of mind" in story comprehension. *Cognition*, *57*, 109–128.

Floreano, D., & Mattiussi, C. (2008). *Bio-inspired artificial intelligence*. Cambridge: MIT Press.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge: MIT Press.

Fogassi, L., Ferrari, P., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: From action organization to intention understanding. *Science*, *308*(5722), 662–667.

Forbus, K., Gentner, D., & Law, K. (1995). MAC/FAC: A model of similarity-based retrieval. *Cognitive Science*, *19*, 144–205.

Forgas, J. P. (1995). Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin*, *11*, 39–66.

Forsterling, F. (2001). *Attribution: An introduction to theories, research, and applications*. New York, NY: Psychology Press.

Fox, E. (2008). *Emotion science*. New York, NY: Palgrave Macmillan.

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition & Emotion*, *14*, 61–92.

Frank, L. G. (1986). Social organization of the spotted hyaena (*Crocuta crocuta*): II. Dominance and reproduction. *Animal Behavior*, *34*, 1510–1527.

Frassinetti, F., Pavani, F., & Làdavas, E. (2002). Acoustical vision of neglected stimuli: Interaction among spatially converging audio-visual inputs in neglect patients. *Journal of Cognitive Neuroscience*, *14*, 62–69.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210.

Fredrickson, B. L. (1998). What good are positive emotions? *Review of General Psychology*, *2*, 300–319.

Freedman, D. (1994). *Brainmakers*. New York, NY: Simon & Schuster.

Freeman, J. B., Stolier, R. M., Ingbretsen, Z. A., & Hehman, E. A. (2014). Amygdala responsivity to high-level social information from unseen faces. *Journal of Neuroscience*, *34*(32), 10573–10581.

French, R. M. (1992). Semi-distributed representations and catastrophic forgetting in connectionist networks. *Connection Science*, *4*, 365–377.

French, R. M. (1999). Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, *3*(4), 128–135.

Fridovich-Kell, J. L., & Diaz, J. M. (2020). Genetically modified organism. In *Encyclopaedia Britannica*.

Friedenberg, J. (2008). *Artificial psychology: The quest for what it means to be human*. New York, NY: Psychology Press.

Friedenberg, J. (2009). *Dynamical psychology: Complexity, self-organization, and mind*. Litchfield Park, AZ: Emergent.

Frieze, I. H., Parsons, J. E., Johnson, P. B., Ruble, D. N., & Zellman, G. L. (1978). *Women and sex roles: A social psychological perspective*. New York, NY: W. W. Norton.

Friston, K. J. (2019). Waves of prediction. *PLoS Biology*, *17*(10), e3000426.

Frith, C. D., & Frith, U. (1999). Interacting minds: A biological basis. *Science*, *286*(5445), 1692–1695.

Frith, U. (2001). What framework should we use for understanding developmental disorders? *Developmental Neuropsychology*, *20*(2), 555–563.

Fromkin, V., Krashen, S., Curtiss, S., Rigler, D., & Rigler, M. (1974). The development of language in Genie: A case of language acquisition beyond the "critical period." *Brain & Language*, *1*(1), 81–107.

Fusaroli, R., Bahrami, B., Olsen, K., Roepstorff, A., Rees, G., Frith, C., & Tylén, K. (2012). Coming to terms: Quantifying the benefits of linguistic coordination. *Psychological Science*, *23*(8), 931–939.

Fuster, J. (2015). *The prefrontal cortex*. New York, NY: Academic Press.

Gagnon, S., & Dore, F. Y. (1994). Cross-sectional study of object permanence in domestic puppies (*Canis familiaris*). *Journal of Comparative Psychology*, *108*, 220–232.

Gallagher, H. L. (2000). Reading the mind in cartoons and stories: An fMRI study of "theory of mind" in verbal and nonverbal tasks. *Neuropsychologia*, *38*(1), 11–21.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of "theory of mind." *Trends in Cognitive Sciences*, 7, 77–83.

Gallup, G. G., Anderson, R., & Shilito, D. J. (2002). The mirror test. In M. Bekoff, C. Allen, & G. M. Burghardt (Eds.), *The cognitive animal* (pp. 325–333). Cambridge: MIT Press.

Gandhi, S. P., Heeger, D. J., & Boynton, G. M. (1999). Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, *96*(6), 3314–3319.

Gantz, B. G., & Turner, C. W. (2004). Combining acoustic and electrical speech processing: Iowa/Nucleus hybrid implant. *Acta Otolaryngol*, *124*, 344–347.

Garcez, A., Lamb, L., Gabbay, D. (2009). *Neural-symbolic cognitive reasoning*. New York, NY: Springer.

Gardner, H. (1974). *The shattered mind*. New York, NY: Vintage Books.

Gardner, H. (1985). *The mind's new science*. New York, NY: Basic Books.

Gardner, H. (2011). *Frames of mind: The theory of multiple intelligences* (3rd ed.). New York, NY: Basic Books.

Gardner, R. A., Gardner, B. T., & Van Cantfort, T. E. (1989). *Teaching sign language to chimpanzees*. New York: State University of New York Press.

Garfield, J. L. (1995). Philosophy: Foundations of cognitive science. In N. A. Stillings, S. E. Weisler, C. H. Chase, M. H. Feinstein, J. L. Garfield, & E. L. Rissland (Eds.), *Cognitive science: An introduction* (2nd ed., pp. 331–377). Cambridge: MIT Press.

Gaulin, S. J. C., & McBurney, D. H. (2001). *Psychology: An evolutionary approach*. Upper Saddle River, NJ: Prentice Hall.

Gaulin, S., & FitzGerald, R. (1986). Sex differences in spatial ability: An evolutionary hypothesis and test. *American Naturalist*, *127*, 74–88.

Gaulin, S., & FitzGerald, R. (1989). Sexual selection for spatial learning ability. *Animal Behavior*, *37*(2), 322–331.

Gawronski, B. (2003). Implicational schemata and the correspondence bias: On the diagnostic value of situationally constrained behavior. *Journal of Personality and Social Psychology*, *84*(6), 1154–1171.

Gazzaniga, M. S. (2005). Forty-five years of split-brain research and still going strong. *Nature Reviews Neuroscience*, *6*(8), 653–659.

Ge, S. S., & Wang, C. (2002). Direct adaptive NN control of a class of nonlinear systems. *IEEE Transactions on Neural Networks*, *13*(1), 214–221.

Geschwind, N. (1972). Language and the brain. *Scientific American*, *226*, 76–83.

Getz, L. M., & Toscano, J. C. (2019). Electrophysiological evidence for top-down lexical influences on early speech perception. *Psychological Science*, *30*(6), 830–841.

Gibbs, R. W., Jr. (2005). *Embodiment and cognitive science*. Cambridge, England: Cambridge University Press.

Gibbs, R. W., Jr., & O'Brien, J. E. (1990). Idioms and mental imagery: The metaphorical motivation for idiomatic meaning. *Cognition*, *36*(1), 35–68.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston, MA: Houghton-Mifflin.

Giedd, J. N., Blumenthal, J., Molloy, E., & Castellanos, F. X. (2001). Brain imaging of attention-deficit/hyperactivity disorder. *Annals of the New York Academy of the Sciences*, *931*, 33–49.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction. *Psychological Review*, *102*, 684–704.

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition*, *43*, 127–171.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences*, *103*(2), 489–494.

Gilbert, D. T. (1998). Ordinary personality. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, 4th ed., pp. 89–150). New York, NY: McGraw-Hill.

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychological Review*, *108*(1), 33–56.

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*(3), 379–401.

Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the revolution to embodiment: 25 years of cognitive psychology. *Perspectives on Psychological Science*, *8*(5), 573–585.

Glushko, R. J. (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *5*(4), 674–691.

Gödel, K. (1934). *On undecidable propositions of formal mathematical systems*. New York, NY: Raven Press.

Goertzel, B. (2016). Infusing advanced AGIs with human-like value systems: Two theses. *Journal of Evolution and Technology*, *26*(1), 50–72.

Goertzel, B., Orseau, L., & Snaider, J. (2015). Artificial general intelligence. *Scholarpedia*, *10*(11), 31847.

Gold, E. M. (1967). Language identification in the limit. *Information and Control*, *10*(5), 447–474.

Goldenberg, G., Podreka, I., Steiner, M., & Willmes, K. (1989). Regional cerebral blood flow patterns in visual imagery. *Neuropsychologia*, *27*(5), 641–664.

Goldinger, S. D., Papesh, M. H., Barnhart, A. S., Hansen, W. A., & Hout, M. C. (2016). The poverty of embodied cognition. *Psychonomic Bulletin & Review*, *23*(4), 959–978.

Goldman-Rakic, P. S. (1993). Working memory and the mind. *Scientific American*, *267*, 110–117.

Goldstein, E. B. (2002). *Sensation and perception* (6th ed.). Pacific Grove, CA: Wadsworth-Thompson.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25.

Goodglass, H., & Geschwind, N. (1976). Language disorders (aphasia). In E. C. Carterete & M. P. Friedman (Eds.), *Handbook of perception: Language and speech* (Vol. 7, pp. 389–428). San Diego, CA: Academic Press.

Gould, S. J. (1978). Sociobiology: the art of storytelling. *New Scientist*, *80*(1129), 530–533.

Gould, S. J. (2000). More things in heaven and earth. In H. Rose & S. Rose (Eds.), *Arguments against evolutionary psychology* (pp. 101–126). New York, NY: Harmony Books.

Gow, D. W., Jr., & Olson, B. B. (2016). Sentential influences on acoustic-phonetic processing: A Granger causality analysis of multimodal imaging data. *Language, Cognition and Neuroscience*, *31*(7), 841–855.

Grafton, S. T., Mazziotta, J. C., Presty, S., Friston, K. J., Frackowiak, R. S., & Phelps, M. E. (1992). Functional anatomy of human procedural learning determined with regional cerebral blood flow and PET. *Journal of Neuroscience*, *12*(7), 2542–2548.

Grafton, S. T., Woods, R. P., & Tyszka, M. (1994). Functional imaging of procedural motor learning: Relating cerebral blood flow with individual subject performance. *Human Brain Mapping*, *1*(3), 221–234.

Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, *8*, 145–146.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, *14*(5), 462–466.

Gravitz, L. (2011, November 18). An ultrathin brain implant monitors seizures. *MIT Technology Review*. Retrieved from http://www.technologyreview.com/news/426154/an-ultrathin-brain-implant-monitors-seizures/

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit intercolumnar synchronization which reflects global stimulus properties. *Nature*, *338*, 334–337.

Graybiel, A. M., Aosaki, T., Flaherty, A. W., & Kimura, M. (1994). The basal ganglia and adaptive motor control. *Science*, *265*, 1826–1831.

Green, C. D., & Vervaeke, J. (1996). *What kind of explanation, if any, is a connectionist net?* North York, Ontario, Canada: Captus.

Greenfield, P. M., & Savage-Rumbaugh, E. S. (1993). Comparing communicative competence in child and chimp: The pragmatics. *Journal of Child Language*, *20*, 1–26.

Greeno, J. G. (1974). Hobbits and orcs: Acquisition of a sequential concept. *Cognitive Psychology*, *6*, 270–292.

Greenwald, A. G. (1992). New look 3: Unconscious cognition reclaimed. *American Psychologist*, *47*, 766–779.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implication association test. *Journal of Personality and Social Psychology Bulletin*, *74*, 1464–1480.

Grèzes, J., Tucker, M., Armony, J., Ellis, R., & Passingham, R. E. (2003). Objects automatically potentiate action: An fMRI study of implicit processing. *European Journal of Neuroscience*, *17*(12), 2735–2740.

Grossberg, S. (1987). Competitive learning: From interactive activation to adaptive resonance. *Cognitive Science*, *11*(1), 23–63.

Guastello, S. J. (2001). Nonlinear dynamics in psychology. *Discrete Dynamics in Nature and Society*, *6*(1), 11–29.

Guinote, A., & Fiske, S. (2003). Being in the outgroup territory increases stereotypic perceptions of outgroups: Situational sources of category activation. *Group Processes & Intergroup Relations*, *6*(4), 323–331.

Guizzo, E. (2014, July 16). Cynthia Breazeal unveils Jibo, a social robot for the home. *IEEE Spectrum*. Retrieved from http://spectrum.ieee.org/automaton/robotics/home-robots/cynthia-breazeal-unveils-jibo-a-social-robot-for-the-home

Gusnard, D. A., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience*, *2*, 685–694.

Guttman, S., Gilroy, L., & Blake, R. (2007). Spatial grouping in human vision: Temporal structure trumps temporal synchrony. *Vision Research*, *47*, 219–230.

Hall, L., Johansson, P., & Strandberg, T. (2012). Lifting the veil of morality: Choice blindness and attitude reversals on a self-transforming survey. *PLoS One*, *7*(9), e45457.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology*, *54*, 917–924.

Happé, F., Ehlers, S., Fletcher, P., Frith, U., Johansson, M., Gillberg, C., . . . Frith, C. (1996). "Theory of mind" in the brain: Evidence from a PET scan study of Asperger syndrome. *NeuroReport*, *8*(1), 197–201.

Harris-Warrick, R. M., & Marder, E. (1991). Modulation of neural networks for behavior. *Annual Review of Neuroscience*, *14*(1), 39–57.

Harris, L. T., Todorov, A., & Fiske, S. T. (2005). Attributions on the brain: Neuro-imaging dispositional inferences, beyond theory of mind. *NeuroImage*, *28*, 763–769.

Hart, A. J., Whalen, P. J., Shin, L. M., McInerney, S. C., Fischer, H., & Rauch, S. L. (2000). Differential response in the human amygdala to racial outgroup vs. ingroup face stimuli. *NeuroReport*, *11*, 2351–2355.

Hartshorne, C., Weiss, P., & Burks, A. (Eds.). (1931–1958). *Collected papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.

Hawkins, J., & Blakeslee, S. (2004). *On intelligence*. New York, NY: Times Books.

Hayes, J. (1989). *The complete problem solver* (2nd ed.). Hillsdale, NJ: Erlbaum.

Hazy, T. E., Frank, M. J., & O'Reilly, R. C. (2007). Towards an executive without a homunculus: Computational models of the prefrontal cortex/basal ganglia system. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1485), 1601–1613.

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. Oxford, England: Wiley.

Henrich, J., Heine, S. J., & Norenzayan, A. (2009). The weirdest people in the world. *Behavioral and Brain Sciences*, *33*(2–3), 61–83.

Herrmann, E., Call, J., Hernandez-Lloreda, M. V., Hare, B., & Tomasello, M. (2009). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, *317*(5843), 1360–1366.

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*, 177–180.

Hinton, G. E. (1981). A parallel computation that assigns canonical object-based frames of reference. In P. J. Hayes (Ed.), Proceedings of the Seventh Annual International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada (pp. 683–685). San Francisco, CA: Kaufmann.

Hobson, R. (1993). The emotional origins of social understanding. *Philosophical Psychology*, *6*(3), 227–249.

Hochberg, L. R., Bacher, D., Jarosiewicz, B., Masse, N. Y., Simeral, J. D., Vogel, J., . . . Donoghue, J. P. (2012). Reach and grasp by people with tetraplegia using a neutrally controlled robotic arm. *Nature*, *485*(7398), 372–375.

Hockett, C. F. (1960). Logical considerations in the study of animal communication. In W. E. Lanyon & W. N. Tavolga (Eds.), *Animal sounds and communication*. Washington, DC: American Institute of Biological Sciences.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, *117*(4), 500–544.

Holbrook, A. L., Berent, M. K., Krosnick, J. A., Visser, P. S., & Boninger, D. S. (2005). Attitude importance and the accumulation of attitude-relevant knowledge in memory. *Journal of Personality and Social Psychology*, *88*, 749–769.

Holbrook, C., & Fessler, D. M. (2013). Sizing up the threat: The envisioned physical formidability of terrorists tracks their leaders' failures and successes. *Cognition*, *127*(1), 46–56.

Holland, J. H. (1992). Genetic algorithms. *Scientific American*, *267*(1), 66–73.

Holland, J. H. (1998). *Emergence: From chaos to order*. New York, NY: Basic Books.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge: MIT Press.

Howell, S. R., Jankowicz, D., & Becker, S. (2005). A model of grounded language acquisition: Sensorimotor features improve lexical and grammatical learning. *Journal of Memory and Language*, *53*(2), 258–276.

Hubel, D. H. (1982). Exploration of the primary visual cortex, 1955–1978. *Nature*, *299*, 515–524.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, *160*, 106–154.

Hubel, D. H., Wiesel, T. N., & Stryker, M. P. (1978). Anatomical demonstration of orientation columns in macaque monkey. *Journal of Comparative Neurology*, *177*, 361–379.

Huebner, B. (2010). Commonsense concepts of phenomenal consciousness: Does anyone care about functional zombies? *Phenomenology and the Cognitive Sciences*, *9*(1), 133–155.

Huette, S., & Anderson, S. (2012). Negation without symbols: The importance of recurrence and context in linguistic negation. *Journal of Integrative Neuroscience*, *11*(3), 295–312.

Huette, S., & McMurray, B. (2010). Continuous dynamics of color categorization. *Psychonomic Bulletin & Review*, *17*(3), 348–354.

Hugenberg, K., & Bodenhausen, G. (2004). Ambiguity in social categorization: The role of prejudice and facial affect in race categorization. *Psychological Science*, *15*(5), 342–345.

Humphreys, G. W., & Riddoch, M. J. (1987). *To see but not to see: A case study of visual agnosia*. Hillsdale, NJ: Erlbaum.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science*, *286*(5449), 2526–2528.

Jackson, F. (1982). Epiphenomenal qualia. *Philosophical Quarterly*, 32, 127–136.

James, W. (1884). What is an emotion? *Mind*, *9*, 188–205.

James, W. (1890). *The principles of psychology*. New York, NY: Dover.

Johnson, J., & Newport, E. (1989). Critical period effects in second-language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago, IL: University of Chicago Press.

Johnston, L., & Macrae, C. (1994). Changing social stereotypes: The case of the information seeker. *European Journal of Social Psychology*, *24*(5), 581–592.

Johnston, W. A., & Heinz, S. P. (1978). Flexibility and capacity demands of attention. *Journal of Experimental Psychology: General*, *107*, 420–435.

Jones, M. T. (2009). *Artificial intelligence: A systems approach*. Sudbury, MA: Jones & Bartlett.

Jones, P. E. (1995). Contradictions and unanswered questions in the Genie case: A fresh look at the linguistic evidence. *Language & Communication*, *15*(3), 261–280.

Jonides, J., & Smith, E. E. (1997). The architecture of working memory. In M. D. Rugg (Ed.), *Cognitive neuroscience* (pp. 243–272). Cambridge: MIT Press.

Jonides, J., Smith, E. E., Koeppe, R. A., Awh, E., Minoshima, S., & Mintun, M. A. (1993). Spatial working memory in humans as revealed by PET. *Nature*, *363*, 623–625.

Judd, C. M., Park, B., Yzerbyt, V. Y., Gordijn, E. H., & Muller, D. (2005). Attributions of intergroup bias and

outgroup homogeneity to ingroup and outgroup others. *European Journal of Social Psychology*, *35*, 677–704.

Kaas, J. H. (2000). Why is the brain size so important: Design problems and solutions as neocortex gets bigger or smaller. *Brain and Mind*, *1*, 7–23.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

Kalat, J. W. (2004). *Biological psychology* (8th ed.). Belmont, CA: Wadsworth/Thomson.

Kamil, A. C. (1978). Systematic foraging by nectar-feeding bird, the amakihi (*Loxops virens*). *Journal of Comparative & Physiological Psychology*, *92*, 388–398.

Kan, I. P., & Thompson-Schill, S. L. (2004). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(1), 43–57.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, *17*(11), 4302–4311.

Karmiloff-Smith, A. (1993). *Beyond modularity: A developmental perspective on cognitive science*. Cambridge: MIT Press.

Karmiloff-Smith, A. (2000). Why babies' brains are not Swiss army knives. In H. Rose & S. Rose (Eds.), *Arguments against evolutionary psychology* (pp. 173–188). New York, NY: Harmony Books.

Karnath, H. O., Ferber, S., & Himmelbach, M. (2001). Spatial awareness is a function of the temporal, not the posterior parietal lobe. *Nature*, *411*(6840), 950–953.

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, *38*(7), 1033–1050.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language*, *32*(4), 474–516.

Kayrak, T. (n.d.). A* algorithm for path planning of robotic soccer player. Retrieved from http://www.eng.auburn.edu/~troppel/courses/5530%202011C%20Robots%20Fall%2011/projects/project%20submissions/written21.pdf

Kehoe, B., Patil, S., Abbeel, P., & Goldberg, K. (Eds.). (2015). Cloud robotics and automation [Special issue]. *IEEE Transactions on Automation Science and Engineering (T-ASE)*, *12*(2).

Keil, F. C. (1989). *Concepts, kinds, and conceptual development*. Cambridge: MIT Press.

Kelley, H. (1950). The warm–cold variable in first impressions of persons. *Journal of Personality*, *18*, 431–439.

Kelley, H. H. (1972). Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weinder (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 1–26). Morristown, NJ: General Learning Press.

Kello, C. T., Brown, G., Ferrer-i-Cancho, R., Holden, J., Linkenkaer-Hansen, K., & Rhodes, T. (2010). Scaling laws in cognitive sciences. *Trends in Cognitive Sciences*, *14*(5), 223–232.

Kello, C. T., Anderson, G. G., Holden, J. G., & Van Orden, G. C. (2008). The pervasiveness of 1/f scaling in speech reflects the metastable basis of cognition. *Cognitive Science*, *32*(7), 1217–1231.

Kelso, J. A. (1995). *Dynamic patterns: The self-organization of brain and behavior*. Cambridge: MIT Press.

Keltner, D., & Shiota, M. N. (2003). New displays and new emotions: A commentary on Rozin and Cohen. *Emotion*, *3*(1), 86–91.

Kerster, B. E., Rhodes, T., & Kello, C. T. (2016). Spatial memory in foraging games. *Cognition*, *148*, 85–96.

Keysers, C., Xiao, D. K., Foldiak, P., & Perrett, D. I. (2005). Out of sight but not out of mind: The neurophysiology of iconic memory in the superior temporal sulcus. *Cognitive Neuropsychology*, *22*(3–4), 316–332.

Kiefer, M., & Pulvermüller, F. (2012). Conceptual representations in mind and brain: Theoretical developments, current evidence and future directions. *Cortex*, *48*(7), 805–825.

Kingsbury, M., & Finlay, B. (2001). The cortex in multidimensional space: Where do cortical areas come from? *Developmental Science*, *48*, 135–152.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence*, *73*(1–2), 31–68.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science*, *18*(4), 513–549.

Kleene, S. C. (1935). A theory of positive integers in formal logic. *American Journal of Mathematics*, *57*, 153–173, 219–244.

Klin, A. (1991). Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*, *21*(1), 29–42.

Knobe, J. (2003). Intentional action and side effects in ordinary language. *Analysis*, *63*, 190–193.

Koch, C. (2004). *The quest for consciousness: A neurobiological approach*. Englewood, CO: Roberts.

Kohler, W. (1920). *Static and stationary physical gestalts*. Braunschweig, Germany: Vieweg.

Kohler, W. (1927). *The mentality of apes*. New York, NY: Harcourt Brace.

Kortge, C. (1990). Episodic memory in connectionist networks. In Proceedings of the Twelfth Annual Conference of the Cognitive Science Society (pp. 764–771). Hillsdale, NJ: Erlbaum.

Kosslyn, S. M. (1975). Information representation in visual images. *Cognitive Psychology*, 7, 341–370.

Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S. M., & Schwartz, S. P. (1977). A simulation of visual imagery. *Cognitive Science*, *1*, 265–295.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 47–60.

Kosslyn, S. M., Thompson, W. L., Kim, I. J., & Alpert, N. M. (1995). Topographical representations of mental images in primary visual cortex. *Nature*, *378*(6556), 496–498.

Koza, J. R. (1992). *Genetic programming: On the programming of computers by means of natural selection*. Cambridge: MIT Press.

Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin*, *21*, 58–75.

Kremer, S. C. (1995). On the computational powers of Elman-style recurrent networks. *IEEE Transactions on Neural Networks*, *6*, 1000–1004.

Kriegman, S., Blackiston, D., Levin, M., & Bongard, J. (2020). A scalable pipeline for designing reconfigurable organisms. *Proceedings of the National Academy of Sciences*, *117*(4), 1853–1859.

Kruse, R., Borgelt, C., Klawonn, F., Moewes, C., Steinbrecher, M., & Held, P. (2013). *Computational intelligence: A methodological introduction*. London, England: Springer-Verlag.

Kubovy, M., & Gepshtein, S. (2003). Perceptual grouping in space and in space-time: An exercise in phenomenological psychophysics. In R. Kimchi, M. Behrmann, & C. R. Olson (Eds.), *Perceptual organization in vision: Behavioral and neural perspectives* (pp. 45–86). Mahwah, NJ: Erlbaum.

Kubovy, M., & Holcombe, A. O. (1998). On the lawfulness of grouping by proximity. *Cognitive Psychology*, *35*(1), 71–98.

Kubovy, M., & Van den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review*, *115*(1), 131–154.

Kubovy, M., & Wagemans, J. (1995). Grouping by proximity and multistability in dot lattices: A quantitative Gestalt theory. *Psychological Science*, *6*(4), 225–234.

Kuczaj, S. A. (1978). Children's judgments of grammatical and ungrammatical past tense verbs. *Child Development*, 49, 319–326.

Kuhlen, A. K., Allefeld, C., & Haynes, J. D. (2012). Content-specific coordination of listeners' to speakers' EEG during communication. *Frontiers in Human Neuroscience*, *6*, 266.

Kulik, L. (2005). Intrafamiliar congruence in gender role attitudes and ethnic stereotypes: The Israeli case. *Journal of Comparative Family Studies*, *36*(2), 289–305.

Kuramoto, Y. (1984). *Chemical oscillations, waves, and turbulence*. Berlin, Germany: Springer-Verlag.

Kurzweil, R. (2005). *The singularity is near: When humans transcend biology*. New York, NY: Penguin.

Kurzweil, R. (2012). *How to create a mind: The secret of human thought revealed*. London, England: Penguin Books.

Kveraga, K., Boshyan, J., & Bar, M. (2007). Magnocellular projections as the trigger of top-down facilitation in recognition. *Journal of Neuroscience*, *27*(48), 13232–13240.

LaBerge, D., & Buchsbaum, M. S. (1990). Positron emission tomographic measurements of pulvinar activity during an attention task. *Journal of Neuroscience*, *10*, 613–619.

Lago-Fernandez, L. F., Huerta, R., Corbacho, F., & Siguenza, J. A. (2000). Fast response and temporal coherent oscillations in small-world networks. *Physical Review Letters*, *84*(12), 2758–2761.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: University of Chicago Press.

Lakoff, G. (2008). *The political mind: Why you cannot understand 21st-century American politics with an 18th-century brain*. New York, NY: Viking Press.

Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Lakoff, G., & Kovecses, Z. (1987). The cognitive model of anger inherent in American English. In N. Quinn & D. Holland (Eds.), *Cultural models in language and thought* (pp. 195–221). New York, NY: Cambridge University Press.

Landauer, T. K. (1986). How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cognitive Science*, *10*(4), 477–493.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Lane, H. (1976). *The wild boy of Aveyron*. Cambridge, MA: Harvard University Press.

Lang, P. J. (1994). The motivational organization of emotion: Affect-reflex connections. In H. M. Stephanie, N. E. van de Poll, & J. A. Sergeant (Eds.), *Emotions: Essays on emotion theory* (pp. 61–93). Hillsdale, NJ: Erlbaum.

Langton, C. (Ed.). (1989). *Artificial life*. Redwood City, CA: Addison-Wesley.

Laplace, P. S. (1902). *A philosophical essay on probabilities*. New York, NY: Wiley.

Lars, S., & Kubovy, M. (2006). On the surprising salience of curvature in grouping by proximity. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 226–234.

Lashley, K. (1950). In search of the engram. *Symposia of the Society of Experimental Biology*, *4*, 454–482.

Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, *87*, 198701.

LeDoux, J. E. (1996). *The emotional brain: The mysterious underpinnings of emotional life*. New York, NY: Simon & Schuster.

Lerner, M. J. (1980). *The belief in a just world: A fundamental delusion*. New York, NY: Plenum Press.

Levine, D. N., & Calvanio, R. (1989). Prosopagnosia: A defect in visual configural processing. *Brain and Cognition*, *10*, 149–170.

Levine, J. (1983). Materialism and qualia: The explanatory gap. *Pacific Philosophical Quarterly*, *64*, 354–361.

Levine, L. J., & Pizarro, D. A. (2004). Emotion and memory research: A group overview. *Social Cognition*, *22*, 530–554.

Levy, J., & Heller, W. (1992). Gender differences in human neuropsychological function. In A. A. Gerall, H. Moltz, & I. L. Ward (Eds.), *Handbook of behavioral neurobiology* (pp. 245–274). New York, NY: Plenum Press.

Lieberman, M. D., Gaunt, R., Gilbert, D. T., & Trope, T. (2002). Reflexion and reflection: A social cognitive neuroscience approach to attributional inference. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 34, pp. 100–249). San Diego, CA: Academic Press.

Lieberman, M. D., Hariri, A., & Bookheimer, S. (2001, February 4). Controlling automatic stereotype activation: An fMRI study. Paper presented at the second annual meeting of the Society for Personality and Social Psychology, San Antonio, TX.

Lieberman, M. D., Ochsner, K. N., Gilbert, D. T., & Schacter, D. L. (2001). Do amnesiacs exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychological Science*, *12*, 135–140.

Lindsay, P. H., & Norman, D. A. (1972). *Human information processing: An introduction to psychology*. New York, NY: Academic Press.

Lingle, J. H., Geva, N., Ostrom, T. M., Leippe, M. R., & Baumgardner, M. H. (1979). Thematic effects of person judgments on impression organization. *Journal of Personality and Social Psychology*, *37*, 674–687.

Liò, P., & Zuliani, P. (Eds.). (2019). *Automated reasoning for systems biology and medicine*. Cham, Switzerland: Springer International Publishing.

Llinas, R. R. (2002). *I of the vortex: From neurons to self*. Cambridge: MIT Press.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, *127*(2), 267–286.

Loftus, E. F. (1979). The malleability of human memory. *American Scientist*, *67*(3), 312–320.

Loftus, E. F. (1995). Remembering dangerously. *Skeptical Inquirer*, *19*, 20–29.

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behaviour*, *13*, 585–589.

Logie, R. H., Zucco, G., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica*, *75*, 55–74.

Losch, M. E., & Cacioppo, J. T. (1990). Cognitive dissonance may enhance sympathetic tonus, but attitudes are changed to reduce negative affect rather than arousal. *Journal of Experimental Social Psychology*, *26*, 289–304.

Louwerse, M. M. (2011). Symbol interdependency in symbolic and embodied cognition. *Topics in Cognitive Science*, *3*(2), 273–302.

Louwerse, M. M. (2018). Knowing the meaning of a word by the linguistic and perceptual company it keeps. *Topics in Cognitive Science*, *10*(3), 573–589.

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science*, *36*(8), 1404–1426.

Louwerse, M. M., Hutchinson, S., Tillman, R., & Recchia, G. (2015). Effect size matters: The role of language statistics and perceptual simulation in conceptual processing. *Language, Cognition and Neuroscience*, *30*(4), 430–447.

Lu, Z., & Sperling, G. (2003). Measuring sensory memory: Magnetoencephalography habituation and psychophysics. In Z. Lu & L. Kaufman (Eds.), *Magnetic source imaging of the human brain* (pp. 319–342). Mahwah, NJ: Erlbaum.

Luck, S. J., & Hillyard, S. A. (1994). Spatial filtering during visual search: Evidence from human electrophysiology. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 1000–1014.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208.

Lupyan, G., Mirman, D., Hamilton, R., & Thompson-Schill, S. L. (2012). Categorization is modulated by transcranial direct current stimulation over left prefrontal cortex. *Cognition*, *124*(1), 36–49.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203.

Maddox, K. B., & Chase, S. G. (2004). Manipulating subcategory salience: Exploring the link between skin tone and social perception of Blacks. *European Journal of Social Psychology*, *34*, 533–546.

Maestrini, E., Paul, A., Monaco, A. P., & Bailey, A. (2000). Identifying autism susceptibility genes. *Neuron*, *28*(1), 19–24.

Mahon, B. Z. (2015). What is embodied about cognition? *Language, Cognition and Neuroscience*, *30*(4), 420–429.

Mahon, B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology–Paris*, *102*(1–3), 59–70.

Malik, H. (Ed.). (2020). *Metaheuristic and evolutionary computation*. Singapore: Springer.

Maloney, J. C. (1999). Functionalism. In *MIT encyclopedia of the cognitive sciences* (pp. 332–334). Cambridge: MIT Press.

Malpass, R. S. (1969). Effects of attitude on learning and memory: The influence of instruction-induced sets. *Journal of Experimental Social Psychology*, *5*, 441–453.

Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, *23*(2), 250–269.

Mandler, J. M. (1992). How to build a baby: II. Conceptual primitives. *Psychological Review*, *99*(4), 587–604.

Manson, J., & Wrangham, R. W. (1991). Intergroup aggression in chimpanzees and humans. *Current Anthropology*, *32*, 369–390.

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Sciences*, *8*(2), 79–86.

Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition*, *46*(1), 53–85.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, J., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development*, *57*, 1–178.

Markoff, J. (2010, September 4). The boss is robotic, and rolling up behind you. *New York Times*. Retrieved from http://www.nytimes.com/2010/09/05/science/05robots.html

Markoff, J. (2014, August 7). IBM develops a new chip that functions like a brain. *New York Times*. Retrieved from http://www.nytimes.com/2014/08/08/science/new-computer-chip-is-designed-to-work-like-the-brain.html?_r=0

Markoff, J. (2014, September 1). Brainy, yes, but far from handy: Building a robot with human touch. *New York Times*. Retrieved from http://www.nytimes.com/2014/09/02/science/robot-touch.html?_r=0

Marks, I. M. (1987). *Fears, phobias, and rituals*. New York, NY: Oxford University Press.

Marks, I. M., & Nesse, R. M. (1994). Fear and fitness: An evolutionary analysis of anxiety disorders. *Ethology and Sociobiology*, *15*, 247–261.

Marler, P. (1970). A comparative approach to vocal learning: Song development in white-crowned sparrows. *Journal of Comparative Physiology and Psychology*, *71*, 1–25.

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London B*, *200*, 269–294.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102.

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45.

Marx, M., & Hillix, W. A. (1979). *Systems and theories in psychology* (3rd ed.). New York, NY: McGraw-Hill.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge: MIT Press.

Mateos, L. A. (2020). Particle robots: A new species of hybrid bio-inspired robotics. *arXiv preprint arXiv:2003.08289*.

Matlock, T. (2004). Fictive motion as cognitive simulation. *Memory & Cognition*, *32*(8), 1389–1400.

Matlock, T., Sparks, D., Matthews, J. L., Hunter, J., & Huette, S. (2012). Smashing new results on aspectual framing: How people talk about car accidents. *Studies in Language*, *36*(3), 699–720.

Matsumoto, R., Nair, D. R., LaPresto, E., Najm, I., Bingaman, W., Shibasaki, H., & Lüders, H. O. (2004). Functional connectivity in the human language system: A cortico-cortical evoked potential study. *Brain*, *127*(10), 2316–2330.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review*, *86*(4), 287–330.

McClelland, J. L. (2009). The place of modeling in cognitive science. *Topics in Cognitive Science*, *1*(1), 11–38.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, *4*(4), 310–322.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–457.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences*, *10*(8), 363–369.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well-defined or fuzzy sets? *Memory & Cognition*, *6*(4), 462–472.

McComb, K. C., Packer, C., & Pusey, A. (1994). Roaring and numerical assessment in contests between groups of female lions, *Panthera leo*. *Animal Behaviour*, *47*, 379–387.

McCorduck, P. (2004). *Machines who think* (2nd ed.). Natick, MA: A. K. Peters.

McCormick, B. H., & Mayerich, D. M. (2004). Three-dimensional imaging using knife-edge scanning microscopy. *Microscopy & Microanalysis*, *10*(2), 1466–1467.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, *5*, 115–133.

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience*, *27*, 1–28.

McGonigle, B. O., & Chalmers, M. (1977). Are monkeys logical? *Nature*, *267*, 694–946.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748.

McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, *21*, 289–311.

McNamara, T. P. (1992). Priming and constraints in places on theories of memory and retrieval. *Psychological Review*, *99*(4), 650–662.

McRae, K., & Hetherington, P. (1993). Catastrophic interference is eliminated in pretrained networks. In Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society (pp. 723–728). Hillsdale, NJ: Erlbaum.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, *126*(2), 99–130.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283–312.

Merolla, P. A., Arthur, J. V., Alvarez-Icaza, R., Cassidy, A. S., Sawada, J., Akopyan, F., . . . Modha, D. S. (2014). A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science*, *345*(6197), 668–673.

Mesulam, M.-M. (1981). A cortical network for directed attention and unilateral neglect. *Annals of Neurology*, *10*, 309–325.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*, 227–234.

Milgram, S. (1967). The small world problem. *Psychology Today*, *2*, 60–67.

Milham, M. P., Banich, M. T., Webb, A., Barad, V., Cohen, N. J., Wszalek, T., & Kramer, A. F. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research*, *12*, 467–473.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*, 81–97.

Miller, R. A., Pople, H. E., Jr., & Myers, J. D. (1982). INTERNIST-1: An experimental computer-based diagnostic consultant for general internal medicine. *New England Journal of Medicine*, *307*, 468–476.

Mills, D. L., Coffey-Corina, S. A., & Neville, H. J. (1993). Language acquisition and cerebral specialization in twenty-month-old infants. *Journal of Cognitive Neuroscience*, *5*(3), 317–334.

Miltenberger, R. G. (2008). Teaching safety skills to children: Prevention of firearm injury as an exemplar of best practice in assessment, training, and generalization of safety skills. *Behavior Analysis in Practice*, *1*(1), 30–36.

Minsky, M. (1975). A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York, NY: McGraw-Hill.

Minsky, M., & Papert, S. (1969). *Perceptrons*. Cambridge: MIT Press.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science*, *26*, 393–424.

Mitchell, J. P. (2008). Contributions of functional neuroimaging to the study of social cognition. *Current Directions in Psychological Science*, *17*(2), 142–146.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2004). Encoding specific effects of social cognition on the neural correlates of subsequent memory. *Journal of Neuroscience*, *24*, 4912–4917.

Miura, I. T., Okamoto, Y., Kim, C. C., & Steere, M. (1993). First graders' cognitive representation of number and understanding of place value: Cross-national comparisons: France, Japan, Korea, Sweden, and the United States. *Journal of Educational Psychology*, *85*(1), 24–30.

Mohr, J. P. (1976). Broca's area and Broca's aphasia. In H. Whitaker & H. Whitaker (Eds.), *Studies in neurolinguistics* (pp. 201–233). New York, NY: Academic Press.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*. Cambridge: MIT Press.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology*, *11*, 56–60.

Morris, J. S., Öhman, A., & Dolan, R. J. (1998). Conscious and unconscious emotional learning in the human amygdala. *Nature*, *393*, 467–470.

Morris, R. G., Ahmed, S., Syed, G. M., & Toone, B. K. (1993). Neural correlates of planning ability: Frontal lobe activation during the Tower of London test. *Neuropsychologia*, *31*, 1367–1378.

Mountcastle, V. (1997). The columnar organization of the neocortex. *Brain*, *120*, 701–722.

Mountcastle, V. B. (1998). *Perceptual neuroscience: The cerebral cortex*. Cambridge, MA: Harvard University Press.

Mundy, P., Card, J., & Fox, N. (2000). Fourteen-month cortical activity and different infant joint attention skills. *Developmental Psychobiology*, *36*, 325–338.

Mundy, P., & Newell, L. (2007). Attention, joint attention, and social cognition. *Current Directions in Psychological Science*, *16*(5), 269–274.

Musallam, S., Corneil, B. D., Greger, B., Scherberger, H., & Andersen, R. A. (2004). Cognitive signals for neural prosthetics. *Science*, *305*, 162–163.

Myers, D. G. (2001). *Psychology*. New York, NY: Worth.

Myerson, R. B. (1997). *Game theory: Analysis of conflict*. Cambridge, MA: Harvard University Press.

Nagel, T. (1974). What is it like to be a bat? *Philosophical Review*, *83*, 435–450.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature*, *320*(6059), 264–265.

Nazir, T. A., Boulenger, V., Roy, A., Silber, B., Jeannerod, M., & Paulignan, Y. (2008). Language-induced motor perturbations during the execution of a reaching movement. *Quarterly Journal of Experimental Psychology*, *61*(6), 933–943.

Néda, Z., Ravasz, E., Brechet, Y., Vicsek, T., & Barabási, A. L. (2000). The sound of many hands clapping. *Nature*, *403*(6772), 849–850.

Neisser, U. (1967). *Cognitive psychology*. New York, NY: Appleton-Century-Crofts.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. New York, NY: W. H. Freeman.

Neisser, U., & Harsch, N. (1992). Phantom flashbulbs: False recollections of hearing the news about Challenger. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of "flashbulb" memories* (pp. 9–31). New York, NY: Cambridge University Press.

Nelson, E. E., & Panksepp, J. (1998). Brain substrates of infant–mother attachment: Contributions of opioids, oxytocin, and norepinephrine. *Neuroscience and Biobehavioral Reviews*, *22*(3), 437–452.

Nesse, R. M. (1987). An evolutionary perspective on panic disorder and agoraphobia. *Ethology and Sociobiology*, *8*, 735–835.

Nesse, R. M. (2000). Is depression an adaptation? *Archives of General Psychiatry*, *57*, 14–20.

Newell, A. (1991). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM*, *19*(3), 113–126.

Newschaffer, C. J., Croen, L. A., Daniels, J., Giarelli, E., Grether, J. K., Levy, S. E., . . . Windham, G. C. (2007). The epidemiology of autism spectrum disorders. *Annual Review of Public Health*, *28*, 235–258.

Nichols, S., & Knobe, J. (2007). Moral responsibility and determinism: The cognitive science of folk intuitions. *Nous*, *41*, 663–685.

Nilsson, N. J. (1969). *A mobile automaton: An application of artificial intelligence techniques* (Technical Note 40). Menlo Park, CA: AI Center.

Nilsson, N. J. (1998). *Artificial intelligence: A new synthesis*. Burlington, MA: Morgan Kaufmann.

Nilsson, N. J. (2007). The physical symbol system hypothesis: Status and prospects. In M. Lungarella, F. Iida, J. Bongard, & R. Pfeifer (Eds.), *50 Years of artificial intelligence* (Vol. 4850, pp. 9–17). New York, NY: Springer.

Nilsson, N. J. (2010). *The quest for artificial intelligence*. Cambridge, England: Cambridge University Press.

Nisbett, R. (2004). *The geography of thought: How Asians and westerners think differently . . . and why*. New York, NY: Free Press.

Nisbett, R., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice Hall.

Nobre, K., & Kastner, S. (Eds.). (2014). *The Oxford handbook of attention*. Oxford, England: Oxford University Press.

Noelle, D. C. (2012). On the neural basis of rule-guided behavior. *Journal of Integrative Neuroscience*, *11*(4), 453–475.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review*, *75*, 522–536.

Norman, D. A., & Shallice, T. (1980). *Attention to action: Willed and automatic control of behavior* (Technical Report No. 99). La Jolla, CA: Center for Human Information Processing.

Notman, M. T., & Nadelson, C. C. (1991). *Women and men: New perspectives on gender differences*. Washington, DC: American Psychiatric Press.

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, *5*(3), 263–281.

Nudo, R. J., Milliken, G. W., Jenkins, W. M., & Merzenich, M. M. (1996). Use-dependent alterations of movement representations in primary motor cortex of adult squirrel monkeys. *Journal of Neuroscience*, *16*(2), 785–807.

Núñez, R., Allen, M., Gao, R., Rigoli, C. M., Relaford-Doyle, J., & Semenuks, A. (2019). What happened to cognitive science? *Nature Human Behaviour*, *3*(8), 782–791.

O'Reilly, R. C. (2006). Biologically based computational models of high-level cognition. *Science*, *314*(5796), 91–94.

O'Sullivan, C., & Durso, F. (1984). Effect of schema-incongruent information on memory for stereotypical attributes. *Journal of Personality and Social Psychology*, *47*(1), 55–70.

O'Sullivan, M. (2003). The fundamental attribution error in detecting deception: The Boy-Who-Cried-Wolf effect. *Personality and Social Psychology Bulletin*, *29*(10), 1316–1327.

Ochsner, K. N. (2000). Are affective events richly recollected or simply familiar? The experience and process of recognizing feelings past. *Journal of Experimental Psychology: General*, *129*, 242–261.

Ochsner, K. N., & Lieberman, M. D. (2001). The emergence of social cognitive neuroscience. *American Psychologist*, *56*(9), 717–734.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology*, *80*, 381–396.

Olds, E. S., Cowan, W. B., & Jolicœur, P. (2000). Partial orientation pop-out helps difficult search for orientation. *Perception & Psychophysics*, *62*(7), 1341–1347.

Olson, J. M., & Zanna, M. P. (1979). A new look at selective exposure. *Journal of Experimental Social Psychology*, *15*, 1–15.

Ospina, M. B., Krebs Seida, J., Clark, B., Karkhaneh, M., Hartling, L., Tjosvold, L., . . . Smith, V. (2008). Behavioural and developmental interventions for autism spectrum disorder: A clinical systematic review. *PLoS One*, *3*(11), e3755.

Paivio, A. (1990). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.

Palmer, J. A., & Palmer, L. K. (2002). *Evolutionary psychology: The ultimate origins of human behavior*. Boston, MA: Allyn & Bacon.

Panksepp, J. (1993). Neurochemical control of moods and emotions: Amino acids to neuropeptides. In M. Lewis & J. M. Haviland (Eds.), *Handbook of emotions* (pp. 87–107). New York, NY: Guilford Press.

Papesh, M. H. (2015). Just out of reach: On the reliability of the action-sentence compatibility effect. *Journal of Experimental Psychology: General*, *144*(6), e116–e141.

Parkin, A. J. (1996). *Explorations in cognitive neuropsychology*. Oxford, England: Blackwell.

Pärnamets, P., Hall, L., & Johansson, P. (2015). Memory distortions resulting from a choice blindness task. Proceedings of the 37th Annual Conference of the Cognitive Science Society (pp. 1823–1828). Austin, TX: Cognitive Science Society.

Pascual-Leone, A., Tormos, J. M., Keenan, J., Tarazona, F., Canete, C., & Catala, M. D. (1998). Study and modulation of human cortical excitability with transcranial magnetic stimulation. *Journal of Clinical Neuropsychology*, *15*, 333–343.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*(6), 717–733.

Patterson, F. G. (1978). The gesture of a gorilla: Language acquisition in another pongid. *Brain & Language*, *5*(1), 72–97.

Patterson, D. A., & Hennessy, J. L. (2012). *Computer organization and design ARM edition: The hardware software interface*. Cambridge, MA: Morgan Kaufmann.

Perrett, D. I., & Oram, M. W. (1993). Neurophysiology of shape processing. *Image and Visual Computing*, *11*, 317–333.

Perrett, D., Rolls, E. T., & Caan, W. (1982). Visual neurons responsive to faces in the monkey temporal cortex. *Experimental Brain Research*, *47*, 329–342.

Peschl, M. F. (1997). The representational relation between environmental structures and neural systems: Autonomy and environmental dependency in neural knowledge representation. *Nonlinear Dynamics, Psychology, and Life Sciences*, *2*, 99–121.

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology*, *58*, 193–198.

Peterson, S. E., Fox, P. T., Posner, M. I., Mintun, M., & Raichle, M. E. (1988). Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature*, *331*, 585–589.

Petrides, M., Alivisatos, B., Meyer, E., & Evans, A. C. (1993). Functional activation of the human frontal cortex during performance of verbal working memory tasks. *Proceedings of the National Academy of Science USA*, *90*, 878–882.

Petrushin, V. A. (2000). Emotion recognition in speech signal: Experimental study, development, and application. Proceedings of the Sixth International Conference on Spoken Language Processing (pp. 222–225). Beijing, China: International Speech Communication Association.

Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in western Europe. *European Journal of Social Psychology*, *25*, 57–75.

Petty, R. E., Tormala, Z. L., Brinol, P., & Jarvis, W. B. G. (2006). Implicit ambivalence from attitude change: An exploration of the PAST model. *Journal of Personality and Social Psychology*, *90*, 21–41.

Petty, R. E., & Wegener, D. T. (1998). Attitude change: Multiple roles for persuasion variables. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (Vol. 1, 4th ed., pp. 323–390). New York, NY: McGraw-Hill.

Pezzulo, G., Barsalou, L. W., Cangelosi, A., Fischer, M. H., Spivey, M., & McRae, K. (2011). The mechanics of embodiment: A dialog on embodiment and computational modeling. *Frontiers in Psychology*, *2*, 5.

Phelps, E. (2006). Emotion and cognition: Insights from studies of the human amygdala. *Annual Review of Psychology*, *57*, 27–53.

Phelps, E., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological Science*, *17*, 292–299.

Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.). New York, NY: International Universities Press. (Original work published 1938)

Picard, R. W. (1997). *Affective computing*. Cambridge: MIT Press.

Pinker, S. (1997). *How the mind works*. New York, NY: W. W. Norton.

Pinker, S. (2003). *The blank slate: The modern denial of human nature*. New York, NY: Penguin.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multilayered perceptron. *Cognition*, *38*, 43–102.

Plutchik, R. (1980). A general psychoevolutionary theory of emotion. In R. Plutchik & H. Kellerman (Eds.), *Emotion: Theory, research, and experience* (pp. 3–33). New York, NY: Academic Press.

Poling, A., Carr, J. E., & LeBlanc, L. A. (2002). Operant conditioning. In M. Hersen & W. Sledge (Eds.), *Encyclopedia of Psychotherapy* (pp. 272–287). Amsterdam: Elsevier.

Popper, K. R., & Eccles, J. C. (1981). *The self and its brain*. Berlin, Germany: Springer-Verlag.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*, 3–25.

Posner, M. I., & Peterson, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*, 25–42.

Posner, M. I., & Snyder, C. R. R. (1975). Facilitation and inhibition in the processing of signals. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 669–682). New York, NY: Academic Press.

Posner, M. I., Inhoff, A. W., Friedrich, F. J., & Cohen, A. (1987). Isolating attentional systems: A cognitive anatomical analysis. *Psychobiology*, *15*, 107–121.

Posner, M. I., Walker, J. A., Friedrich, F. J., & Rafal, R. D. (1984). Effects of parietal injury on covert orienting of attention. *Journal of Neuroscience*, *4*, 1863–1874.

Powell, M. C., & Fazio, R. H. (1984). Attitude accessibility as a function of repeated attitudinal expression. *Personality and Social Psychology Bulletin*, *10*, 139–148.

Pratto, F., & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology*, *61*, 380–390.

Preissl, R., Theodore, M. W., Datta, P., Flickner, M., Singh, R., Esser, S. K., . . . Modha, D. S. (2012). Compass: A scalable simulator for an architecture for cognitive computing. In Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis (pp. 1–11). Los Alamitos, CA: IEEE Computer Society Press.

Premack, D. (1976). *Intelligence in ape and man*. Mahwah, NJ: Erlbaum.

Prince, A., & Smolensky, P. (1997). Optimality: From neural networks to universal grammar. *Science*, *275*, 1604–1610.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in extrastriate cortex studied by functional MRI. *Journal of Neurophysiology*, *74*, 1192–1199.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793–797.

Purpura, K. P., & Schiff, N. D. (1997). The thalamic intralaminar nuclei: A role in visual awareness. *The Neuroscientist*, *3*, 8–15.

Pylyshyn, Z. W. (1981). The imagery debate: Analog media versus tacit knowledge. In N. Block (Ed.), *Imagery* (pp. 151–206). Cambridge: MIT Press.

Quiroga, R. Q. (2012). Concept cells: the building blocks of declarative memory functions. *Nature Reviews Neuroscience*, *13*(8), 587–597.

Rabbie, J., & Horwitz, M. (1969). Arousal of ingroup-outgroup bias by a chance win or loss. *Journal of Personality and Social Psychology*, *13*, 269–277.

Rafal, R. D., Posner, M. I., Friedman, J. H., Inhoff, A. W., & Bernstein, E. (1988). Orienting of visual attention in progressive supranuclear palsy. *Brain*, *111*, 267–280.

Rakison, D. H., & Derringer, J. (2008). Do infants possess an evolved spider-detection mechanism? *Cognition*, *107*(1), 381–393.

Ramirez, S., Liu, X., Lin, P.-A., Suh, J., Pignatelli, M., Redondo, R. L., . . . Tonegawa, S. (2013). Creating a false memory in the hippocampus. *Science*, *346*(6144), 387–391.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, *2*(1), 79–87.

Ravenscroft, I. (2005). *Philosophy of mind: A beginner's guide*. Oxford, England: Oxford University Press.

Ravizza, S. M., & Ivry, R. B. (2001). Comparison of the basal ganglia and cerebellum in shifting attention. *Journal of Cognitive Neuroscience*, *13*, 285–297.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201.

Reali, F., Spivey, M. J., Tyler, M. J., & Terranova, J. (2006). Inefficient conjunction search made efficient by concurrent spoken delivery of target identity. *Perception & Psychophysics*, *68*(6), 959–974.

Reason, J. T. (1979). Action not as planned. In G. Underwood & R. Stevens (Eds.), *Aspects of consciousness* (pp. 67–89). New York, NY: Academic Press.

Reed, S. K. (2000). *Cognition* (5th ed.). Belmont, CA: Wadsworth Thomson.

Reeke, G. N., Jr., & Edelman, G. M. (1988). Real brains and artificial intelligence. *Daedalus*, *117*(1), 143–173.

Rehder, B., & Kim, S. (2006). How causal knowledge affects classification: A generative theory of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 659–683.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*, 274–280.

Reid, C. R., Latty, T., Dussutour, A., & Beekman, M. (2012). Slime mold uses an externalized spatial "memory" to navigate in complex environments. *Proceedings of the National Academy of Sciences USA*, *109*(43), 17490–17494.

Reisberg, D. (2001). *Cognition: Exploring the science of mind* (2nd ed.). New York, NY: W. W. Norton.

Reisberg, D., Culver, C., Heuer, F., & Fischman, D. (1986). Visual memory: When imagery vividness makes a difference. *Journal of Mental Imagery*, *10*, 51–74.

Reisberg, D., Heuer, F., McLean, J., & O'Shaughnessy, M. (1988). The quantity, not the quality, of affect predicts memory vividness. *Bulletin of the Psychonomic Society*, *26*, 100–103.

Reynolds, C. W. (1987, August). Flocks, herds and schools: A distributed behavioral model. In Proceedings of the 14th annual conference on Computer graphics and interactive techniques (pp. 25–34). New York, NY: Association for Computing Machinery.

Rhodes, T., & Turvey, M. T. (2007). Human memory retrieval as Lévy foraging. *Physica A: Statistical Mechanics and its Applications*, *385*(1), 255–260.

Rhodes, T., Kello, C. T., & Kerster, B. (2014). Intrinsic and extrinsic contributions to heavy tails in visual foraging. *Visual Cognition*, *22*(6), 809–842.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination. *Psychological Science*, *18*(5), 407–413.

Richardson, D. C., Spivey, M. J., Barsalou, L. W., & McRae, K. (2003). Spatial representations activated during real-time comprehension of verbs. *Cognitive Science*, *27*(5), 767–780.

Richardson, D., & Matlock, T. (2007). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition*, *102*(1), 129–138.

Richardson, M. J., & Chemero, A. (2014). Complex dynamical systems and embodiment. In L. Shapiro (Ed.), *The Routledge handbook of embodied cognition* (pp. 39–50). New York, NY: Routledge.

Rilling, J. K., Sanfey, A. G., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2004). The neural correlates of theory of mind within interpersonal interactions. *NeuroImage*, *22*, 1694–1703.

Rinck, M., Glowalia, U., & Schneider, K. (1992). Mood-congruent and mood-incongruent learning. *Memory & Cognition*, *20*, 29–39.

Rips, L. J. (1975). Inductive judgments about natural categories. *Journal of Verbal Learning and Verbal Behavior*, *14*, 665–681.

Rips, L. J., Shoben, E. J., & Smith, E. E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, *12*, 1–20.

Rizzolatti, G., & Craighero, L. (2004). The mirror neuron system. *Annual Review of Neuroscience*, *27*, 169–192.

Robinson, C. (1998). *Dynamical systems: Stability, symbolic dynamics, and chaos*. Boca Raton, FL: CRC Press.

Roediger, H. L. (1980). The effectiveness of four mnemonics in ordering recall. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 558–567.

Rolls, E. T. (1999). *The brain and emotion*. Oxford, England: Oxford University Press.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York, NY: Academic Press.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192–233.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*, 386–408.

Rosenblatt, F. (1959). Perceptron simulation experiments (Project Para). *Proceedings of the IRE*, *48*, 301–309.

Rosenblatt, F. (1962). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. New York, NY: Spartan Books.

Rosset, C. (2020). Turing-nlg: A 17-billion-parameter language model by Microsoft. *Microsoft Research Blog*, *2*, 13.

Rozin, P., Lowery, L., & Ebert, R. (1994). Variety of disgust faces and the structure of disgust. *Journal of Personality and Social Psychology*, *66*(5), 870–881.

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Reviews Neuroscience*, *4*, 546–550.

Ruckmick, C. A. (1913). The use of the term "function" in English textbooks of psychology. *American Journal of Psychology*, *24*, 99–123.

Ruff, C. C., Kristjansson, A., & Driver, J. (2007). Readout from iconic memory and selective spatial attention involves similar neural processes. *Psychological Science*, *18*(10), 901–909.

Rumelhart, D. E., & Todd, P. M. (1993). Learning and connectionist representations. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 3–30). Cambridge: MIT Press.

Rumelhart, D. E., & McClelland, J. L. (1987). Learning the past tenses of English verbs: Implicit rules or parallel distributed processing? In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 195–248). Englewood Cliffs, NJ: Lawrence Erlbaum Associates.

Rumelhart, D. E., McClelland, J. L., & PDP Research Group. (Eds.). (1986). *Parallel distributed processing: Foundations* (Vol. 1). Cambridge: MIT Press.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, *110*, 145–172.

Russell, S. (1979). Brain size and intelligence: A comparative perspective. In D. A. Oakley & H. C. Plotkin (Eds.), *Brain, behavior, and evolution* (pp. 126–153). London, England: Methuen.

Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Rusting, C. L., & DeHart, T. (2000). Retrieving positive memories to regulate negative mood: Consequences for mood-congruent memory. *Journal of Personality and Social Psychology*, *78*, 737–752.

Rusting, C. L., & Nolen-Hoeksema, S. (1998). Regulating responses to anger: Effects of rumination and distraction on angry mood. *Journal of Personality and Social Psychology*, *74*, 790–803.

Rutter, M. (2005). Incidence of autism spectrum disorders: Changes over time and their meaning. *Acta Paediatrica*, *94*(1), 2–15.

Ryle, G. (1949). *The concept of mind*. London, England: Hutchinson.

Sacks, O. (1998). *The man who mistook his wife for a hat*. New York, NY: Touchstone.

Saffran, J. R., & Thiessen, E. D. (2007). Domain general learning capacities. In E. Hoff & M. Schatz (Eds.), *Blackwell handbook of language development* (pp. 68–86). Oxford, England: Blackwell.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Salovey, P. (1990, January–February). Interview. *American Scientist*, pp. 25–29.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, *3*(3), 210–229.

Samuelson, W., & Zeckhauser, R. J. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*, 7–59.

Sapolsky, R. M. (2017). *Behave: The biology of humans at our best and worst*. New York, NY: Penguin.

Sathian, K., Zangaladze, A., Hoffman, J., & Grafton, S. (1997). Feeling with the mind's eye. *NeuroReport*, *8*, 3877–3881.

Savage-Rumbaugh, E. S. (1986). *Ape language: From conditioned response to symbol*. New York, NY: Columbia University Press.

Savage-Rumbaugh, E. S., Murphy, J., Sevcik, R. A., Brakke, K. E., Williams, S., & Rumbaugh, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, 58(Serial No. 233), 3–4.

Saxe, R., & Kanwisher, N. (2003). People thinking about thinking people: fMRI investigations of theory of mind. *NeuroImage*, *19*, 1835–1842.

Scaife, M., & Bruner, J. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*, 265–266.

Scannell, J. W. (1997). Determining cortical landscapes. *Nature*, *386*, 452.

Schachter, S. S., & Singer, J. E. (1962). Cognitive and physiological determinant of emotional state. *Psychological Review*, *69*, 379–399.

Schacter, D., Gilbert, D., Wegner, D., & Hood, B. M. (2011). *Psychology*. London, England: Macmillan International Higher Education.

Schank, R., & Abelson, R. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Hillsdale, NJ: Erlbaum.

Schiller, P. H., Sandell, J. H., & Maunsell, J. H. R. (1987). The effect of frontal eye field and superior colliculus lesions on saccadic latencies in the rhesus monkey. *Journal of Neurophysiology*, *57*, 1033–1049.

Schlenker, B. R. (1980). *Impression management: The self-concept, social identity, and interpersonal relations*. Monterey, CA: Brooks/Cole.

Schultz, D. P., & Schultz, S. E. (1987). *A history of modern psychology* (4th ed.). Orlando, FL: Harcourt Brace Jovanovich.

Schutter, D. (2008). Antidepressant efficacy of high-frequency transcranial magnetic stimulation over the

left dorsolateral prefrontal cortex in double-blind sham-controlled designs: A meta-analysis. *Psychological Medicine*, *39*(1), 65–75.

Schwartz, B. J. (1999). Magnetic source imaging as a clinical tool in functional brain mapping. *GEC Review*, *14*(2), 124–127.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge, England: Cambridge University Press.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*, 417–457.

Searle, J. R. (1984). *Minds, brains, and science*. Cambridge, MA: Harvard University Press.

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge: MIT Press.

Searle, J. R. (2014, October 9). What your computer can't know. *New York Review of Books*. Retrieved from http://www.nybooks.com/articles/archives/2014/oct/09/what-your-computer-cant-know/

Sears, D. O. (1968). The paradox of de facto selective exposure. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook* (pp. 777–787). Chicago, IL: Rand McNally.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: Bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76.

Segal, G. (2000). *A slim book about narrow content*. Cambridge: MIT Press.

Segall, M. H., Campbell, D. T., & Herskovits, M. J. (1966). *The influence of culture on visual perception*. Indianapolis, IN: Bobbs-Merrill.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*(4), 523–568.

Sejnowski, T., & Rosenberg, C. (1986). *NETtalk: A parallel network that learns to read aloud* (Technical Report JHU/EEC-86/01). Baltimore, MD: Johns Hopkins University Press.

Selfridge, O. G. (1959). Pandemonium: A paradigm for learning. In D. Blake & A Uttley (Eds.), Proceedings of the Symposium on the Mechanisation of Thought Processes (pp. 513–526). London, England: H. M. Stationery Office.

Service, R. F. (2014). The brain chip. *Science*, *345*(6197), 614–616.

Shadlen, M. N., & Movshon, J. A. (1999). Synchrony unbound: A critical evaluation of the temporal binding hypothesis. *Neuron*, *24*(1), 67–77.

Shahin, A. J., Backer, K. C., Rosenblum, L. D., & Kerlin, J. R. (2018). Neural mechanisms underlying cross-modal phonetic encoding. *Journal of Neuroscience*, *38*(7), 1835–1849.

Shallice, T. (1982). Specific impairment of planning. *Philosophical Transactions of the Royal Society of London B*, *298*, 199–209.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Shapiro, L. (2019). *Embodied cognition* (2nd ed.). New York, NY: Routledge.

Shebani, Z., & Pulvermüller, F. (2013). Moving the hands and feet specifically impairs working memory for arm- and leg-related action words. *Cortex*, *49*(1), 222–231.

Sheehan, S. (1982). *Is there no place on earth for me?* New York, NY: Vintage Books.

Sheikh, K. (2019, November 19). How the brain can rewire itself after half of it is removed. *New York Times*.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *153*, 652–654.

Shermer, M. (2008). *The mind of the market*. New York, NY: Henry Holt.

Sherry, D., Forbes, M., Khurgel, M., & Ivy, G. (1993). Greater hippocampal size in females of the brood parasitic brown-headed cowbird. *Proceedings of the National Academy of Sciences USA*, *90*, 7839–7843.

Shin, S. S., Dixon, C. E., Okonkwo, D. O., & Richardson, R. M. (2014). Neurostimulation for traumatic brain injury. *Journal of Neurosurgery*, *121*(5), 1219–1231.

Shockley, K., Santana, M. V., & Fowler, C. A. (2003). Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(2), 326–332.

Siegal, M., & Varley, R. (2002). Neural systems involved in theory of mind. *Nature Reviews Neuroscience*, *3*, 463–471.

Siegal, M., Carrington, J., & Radel, M. (1996). Theory of mind and pragmatic understanding following right hemisphere damage. *Brain and Language*, *53*(1), 40–50.

Silverman, G. (2003, March 20–22). The role of neural networks in telerehabilitation. Paper presented at the First International IEEE EMBS Conference on Neural Engineering, Capri, Italy.

Silverman, G. (2006). Computers in the biomedical laboratory. In J. G. Webster (Ed.), *Encyclopedia of medical devices and instrumentation* (2nd ed., pp. 308–315). Hoboken, NJ: Wiley.

Silverman, G., & Brudny, J. (1990, June 17–22). The role of the computer in neuromotor rehabilitation. Paper presented at the VI International Rehabilitation Medicine Association, Madrid, Spain.

Silverman, I., & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 533–549). New York, NY: Oxford University Press.

Silverman, I., & Phillips, K. (1998). The evolutionary psychology of spatial sex differences. In C. Crawford & D. L. Krebs (Eds.), *Handbook of evolutionary psychology: Ideas, issues, and applications* (pp. 595–612). Mahwah, NJ: Erlbaum.

Simmons, W. K., Martin, A., & Barsalou, L. W. (2005). Pictures of appetizing foods activate gustatory cortices for taste and reward. *Cerebral Cortex*, *15*(10), 1602–1608.

Simon, B. (1992). The perception of ingroup and outgroup homogeneity: Reintroducing the intergroup context. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 3, pp. 1–30). Oxford, England: Wiley.

Simon, H. A. (1969). *The sciences of the artificial*. Cambridge: MIT Press.

Simon, L., Greenberg, J., & Brehm, J. (1995). Trivialization: The forgotten mode of dissonance reduction. *Journal of Personality and Social Psychology*, *68*(2), 247–260.

Sims, D. W., Reynolds, A. M., Humphries, N. E., Southall, E. J., Wearmouth, V. J., Metcalfe, B., & Twitchett, R. J. (2014). Hierarchical random walks in trace fossils and the origin of optimal search behavior. *Proceedings of the National Academy of Sciences*, *111*(30), 11073–11078.

Sims, R. W., & Gerard, B. M. (1985). *Earthworms: Keys and notes for the identification and study of the species* (Vol. 31). London, England: Brill Archive.

Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, *303*, 1157–1162.

Singer, W. (1996). Neuronal synchronization: A solution to the binding problem? In R. Llinas & P. S. Churchland (Eds.), *The mind–brain continuum: Sensory processes* (pp. 100–130). Cambridge: MIT Press.

Singer, W. (1999). Binding by neural synchrony. In *MIT encyclopedia of the cognitive sciences* (pp. 81–84). Cambridge: MIT Press.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience*, *18*, 555–586.

Singer, W., Engel, A. K., Kreiter, A. K., Munk, M. H. J., Neuenschwander, S., & Roelfsema, P. R. (1997). Neuronal assemblies: Necessity, signature, and detectability. *Trends in Cognitive Sciences*, *1*(7), 252–261.

Skarda, C. A., & Freeman, W. J. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, *10*(2), 161–173.

Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. Cambridge, MA: B. F. Skinner Foundation.

Sloman, A., Chrisley, R., & Scheutz, M. (2005). The architectural basis of affective states and processes. In J. M. Fellows & M. A. Arbib (Eds.), *Who needs emotions? The brain meets the robot* (pp. 203–244). Oxford, England: Oxford University Press.

Sloman, S. A., Love, B. C., & Ahn, W. K. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*(2), 189–228.

Smaldino, P. E. (2014). The cultural evolution of emergent group-level traits. *Behavioral and Brain Sciences*, *37*(3), 243–295.

Smaldino, P. E., Schank, J. C., & McElreath, R. (2013). Increased costs of cooperation help cooperators in the long run. *The American Naturalist*, *181*(4), 451–463.

Smith, E. E. (2000). Neural bases of human working memory. *Current Directions in Psychological Science*, *9*, 45–49.

Smith, E. E., & Jonides, J. (1994). Working memory in humans: Neuropsychological evidence. In M. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 1009–1020). Cambridge: MIT Press.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, *283*, 1657–1661.

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex*, *6*, 11–20.

Smith, E. E., Jonides, J., Koeppe, R. A., Awh, E., Schumacher, E. H., & Minoshima, S. (1995). Spatial versus object working memory: PET investigations. *Journal of Cognitive Neuroscience*, 7, 337–356.

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*(1), 9–17.

Solomon, R. C. (2003). *What is an emotion? Classic and contemporary readings*. New York, NY: Oxford University Press.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist*, *60*(9), 950–958.

Spencer, J., Thomas, M., McClelland, J. (Eds.). (2009). *Toward a unified theory of development: Connectionism and dynamic systems theory re-considered*. Oxford, England: Oxford University Press.

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74(Whole No. 48), 1–29.

Sperling, G. (1963). A model for visual memory tasks. *Human Factors*, *5*, 9–31.

Sperry, R. W. (1985). The dual brain: Hemispheric specialization in humans. In D. F. Benson & E. Zaidel (Eds.), *UCLA forum in medical sciences* (Vol. 26, pp. 11–26). New York, NY: Guilford Press.

Spivey, M. (2007). *The continuity of mind*. Oxford, England: Oxford University Press.

Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences USA*, *102*(29), 10393–10398.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, *12*(2), 153–156.

Stanley, K. O., & Miikkulainen, R. (2002). Evolving neural networks through augmenting topologies. *Evolutionary Computation*, *10*(2), 99–127.

Steiner, J. E. (1979). Human facial expressions in response to taste and smell stimulation. *Advances in Child Development and Behavior*, *13*, 257–295.

Stepp, N., Chemero, A., & Turvey, M. T. (2011). Philosophy for the rest of cognitive science. *Topics in Cognitive Science*, *3*(2), 425–437.

Sternberg, S. (1966). High-speed scanning in human memory. *Science*, *153*, 652–654.

Sternberg, S. (2016). In defence of high-speed memory scanning. *Quarterly Journal of Experimental Psychology*, *69*(10), 2020–2075.

Stevens, J. A., Fonlupt, P., Shiffrar, M., & Decety, J. (2000). New aspects of motion perception: Selective neural encoding of apparent human movements. *Neuroreport*, *11*(1), 109–115.

Stillings, N. A., Weisler, S. W., Chase, C. H., Feinstein, M. H., Garfield, J. L., & Rissland, E. L. (1995). *Cognitive science: An introduction* (2nd ed.). Cambridge: MIT Press.

Stoean, R., Stoean, C., Becerra-García, R., García-Bermúdez, R., Atencia, M., García-Lagos, F., . . . Joya, G. (2020). A hybrid unsupervised—Deep learning tandem for electrooculography time series analysis. *PLoS One*, *15*(7), e0236401.

Strogatz, S. (2003). *Sync: How order emerges from chaos in the universe, nature, and daily life*. New York, NY: Hyperion.

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes*. New York, NY: Raven Press.

Sutton, R. (1991). *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration*. Cambridge: MIT Press.

Swinney, D. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, *18*(6), 645–659.

Swinney, D., Zurif, E. B., Prather, P., & Love, T. (1996). Neurological distribution of processing resources underlying language comprehension. *Journal of Cognitive Neuroscience*, *8*(2), 174–184.

Tabor, W. (2009). A dynamical systems perspective on the relationship between symbolic and non-symbolic computation. *Cognitive Neurodynamics*, *3*(4), 415–427.

Tabossi, P. (1988). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language*, *27*(3), 324–340.

Talarico, J., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science*, *14*, 455–461.

Talmy, L. (2000). *Toward a cognitive semantics* (Vol. 2). Cambridge: MIT Press.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, *18*(4), 427–440.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Tang-Schomer, M. D., White, J. D., Tien, L. W., Schmitt, L. I., Valentin, T. M., Graziano, D. J., . . . Kaplan, D. L. (2014). Bioengineered functional brain-like cortical tissue. *Proceedings of the National Academy of Sciences USA*, *111*, 13811–13816.

Tang, Y.-P., Shimizu, E., Dube, G. R., Rampon, C., Kerchner, G. A., Zhuo, M., . . . Tsien, J. Z. (1999). Genetic enhancement of learning and memory in mice. *Nature*, *401*, 63–69.

Tarr, M. J., & Gauthier, I. (2000). FFA: A flexible fusiform area for subordinate-level visual processing automatized by expertise. *Nature Neuroscience*, *3*(8), 764–769.

Terrace, H. S., Petitto, L. A., Sanders, R. J., & Bever, T. G. (1979). Can an ape create a sentence? *Science*, *206*(4421), 891–902.

Thagard, P. (2000). *Mind: Introduction to cognitive science*. Cambridge: MIT Press.

Thagard, P. (2008). *Hot thought: Mechanisms and applications of emotional cognition*. Cambridge: MIT Press.

Thelen, E., & Smith, L. B. (1994). *A dynamic systems approach to the development of cognition and action*. Cambridge: MIT Press.

Thomas, L. E., & Irwin, D. E. (2006). Voluntary eyeblinks disrupt iconic memory. *Perception & Psychophysics*, *68*(3), 475–488.

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin & Review*, *14*(4), 663–668.

Thompson-Schill, S. L. (2003). Neuroimaging studies of semantic memory: Inferring "how" from "where." *Neuropsychologia*, *41*(3), 280–292.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, *2*(4), i.

Tillfors, M., Furmark, T., Marteinsdottir, I., Fischer, H., Pissiota, A., Långström, B., & Fredrikson, M. (2001). Cerebral blood flow in subject with social phobia during stressful speaking tasks: A PET study. *American Journal of Psychiatry*, *158*(8), 1220–1226.

Titchener, E. B. (1896). *An outline of psychology*. New York, NY: Macmillan.

Tocco, G., Maren, S., Shors, T. J., Baudry, M., & Thompson, R. F. (1992). Long-term potentiation is associated with increased [3H]AMPA binding in rat hippocampus. *Brain Research*, *573*, 228–234.

Tomasello, M., & Carpenter, M. (2005). The emergence of social cognition in three young chimpanzees. *Monographs of the Society for Research in Child Development*, *70*(1, Serial No. 279), 1–152.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intention: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*, 675–735.

Tomkins, S. S. (1963). *Affect, imagery, and consciousness: The negative affects* (Vol. 2). New York, NY: Springer.

Tong, F., Nakayama, K., Vaughn, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, *21*, 753–759.

Tracy, J. L., & Robins, R. W. (2004). Show your pride: Evidence for a discrete emotion expression. *Psychological Science*, *15*(3), 194–197.

Traut-Mattausch, E., Schulz-Hardt, S., Greitemeyer, T., & Frey, D. (2004). Expectancy confirmation in spite of disconfirming evidence: The case of price increases due to the introduction of the Euro. *European Journal of Social Psychology*, *34*, 739–760.

Treisman, A. (1964). Monitoring and storage of irrelevant messages in selective attention. *Journal of Verbal Learning and Verbal Behavior*, *3*, 449–459.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Trueswell, J. C., & Tanenhaus, M. K. (Eds.). (2005). *Approaches to studying world-situated language use: Bridging*

the language-as-product and language-as-action traditions. Cambridge: MIT Press.

Tsien, J. Z., Huerta, P. T., & Tonegawa, S. (1996). The essential role of hippocampal CA1 NMDA receptor-dependent synaptic plasticity in spatial memory. *Cell*, *87*, 1327–1338.

Tucker-Drob, E. M., & Briley, D. A. (2014). Continuity of genetic and environmental influences on cognition across the life span: A meta-analysis of longitudinal twin and adoption studies. *Psychological Bulletin*, *140*(4), 949–979.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition*, *8*(6), 769–800.

Tucker, M., & Ellis, R. (2004). Action priming by briefly presented objects. *Acta Psychologica*, *116*(2), 185–203.

Turing, A. M. (1937). On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, *2*(1), 230–265.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460.

Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, *9*(5), 160–164.

Turvey, M. T. (2018). *Lectures on perception: An ecological perspective*. New York, NY: Routledge.

Turvey, M. T., & Carello, C. (2012). On intelligence from first principles: Guidelines for inquiry into the hypothesis of physical intelligence (PI). *Ecological Psychology*, *24*(1), 3–32.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.

Tversky, A., & Kahneman, D. (1981). The framing of decision and the psychology of choice. *Science*, *211*, 453–458.

Tyler, L. K. (1992). *Spoken language comprehension: An experimental approach to disordered and normal processing*. Cambridge: MIT Press.

Ulanovsky, N., & Moss, C. F. (2008). What the bat's voice tells the bat's brain. *Proceedings of the National Academy of Sciences USA*, *105*(25), 8491–8498.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge: MIT Press.

Valdez, P., & Mehrabian, A. (1994). Effects of color on emotions. *Journal of Experimental Psychology: General*, *123*(4), 394–409.

Vallone, R. P., Ross, L. D., & Lepper, M. R. (1985). The hostile media phenomenon: Biased perception and perceptions of media bias in coverage of the Beirut massacre. *Journal of Personality and Social Psychology*, *49*, 577–585.

Van den Heuvel, M. P., Bullmore, E. T., & Sporns, O. (2016). Comparative connectomics. *Trends in Cognitive Sciences*, *20*(5), 345–361.

Van Orden, G. C., Holden, J. G., & Turvey, M. T. (2003). Self-organization of cognitive performance. *Journal of Experimental Psychology: General*, *132*(3), 331–350

Van Orden, G. C., & Kloos, H. (2003). The module mistake. *Cortex*, *39*(1), 164–166.

Vanderschraaf, P. (2006). War or peace? A dynamical analysis of anarchy. *Economics and Philosophy*, *22*(2), 243–279.

Vanier, M., & Caplan, D. (1990). CT-scan correlates of agrammatism. In L. Menn & L. Obler (Eds.), *Agrammatic aphasia: A cross-linguistic narrative sourcebook* (pp. 37–114). Amsterdam, Netherlands: John Benjamins.

Varela, F. J., Thompson, E., & Rosch, E. (2017). *The Embodied Mind, revised edition: Cognitive science and human experience*. Cambridge: MIT Press.

Vargas, R., & Just, M. A. (2020). Neural representations of abstract concepts: Identifying underlying neurosemantic dimensions. *Cerebral Cortex*, *30*(4), 2157–2166.

Vargha-Khadem, F., Gadian, D. G., Watkins, K. E., Connelly, A., Van Paesschen, W., & Mishkin, W. (1997). Differential effects of early hippocampal pathology on episodic and semantic memory. *Science*, *277*, 376–380.

Veale, T., & Cardoso, A. F. (2019). *Computational creativity: The philosophy and engineering of autonomously creative systems*. New York, NY: Springer.

Von der Emde, G., Schwarz, S., Gomez, L., Budelli, R., & Grant, K. (1998). Electric fish measure distance in the dark. *Nature*, *395*, 890–894.

von der Malsburg, C. (1981). *The correlation theory of brain function* (Internal Report 81-2). Gottingen, Germany: Max-Planck-Institute for Biophysical Chemistry.

von Fersen, L., Wynne, C. D. L., Delius, J. D., & Staddon, J. E. R. (1991). Transitive inference formation in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *17*, 334–341.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, *7*, 270–276.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences*, *9*, 585–594.

Vuilleumier, P., & Schwartz, S. (2001). Emotional facial expressions capture attention. *Neurology*, *56*(2), 153–158.

Wachsmuth, E., Oram, M. W., & Perrett, D. I. (1994). Recognition of objects and their component parts: Responses of single units in the temporal cortex of the macaque. *Cerebral Cortex*, *4*, 509–522.

Wallas, E. (1926). *The art of thought*. New York, NY: Harcourt Brace.

Wallraff, H. G. (1980). Olfaction and homing pigeons: Nerve-section experiments, critique, hypotheses. *Journal of Comparative Physiology*, *139*, 209–224.

Walsh, V., & Pascual-Leone, A. (2005). *Transcranial magnetic stimulation: A neurochronometrics of mind*. Cambridge: MIT Press.

Wang, P., & Hammer, P. (2015). Assumptions of decision-making models in AGI. In J. Bieger, B. Goertzel, & A. Potapov (Eds.), *Artificial general intelligence: AGI 2015* (pp. 197–207). Cham, Switzerland: Springer.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392–393.

Wasserman, E. A., & Zentall, T. R. (2006). *Comparative cognition*. Oxford, England: Oxford University Press.

Watts, D. J. (2003). *Six degrees: The science of a connected age*. New York, NY: W. W. Norton.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature*, *393*, 440–442.

Weinberg, J., Nichols, S., & Stich, S. (2001). Normative and epistemic intuitions. *Philosophical Topics*, *29*, 429–460.

Weizenbaum, H. (1976). *Computer power and human reason*. San Francisco, CA: W. H. Freeman.

Wender, P. H., Wolf, L. E., & Wasserstein, J. (2001). Adults with ADHD: An overview. *Annals of the New York Academy of Sciences*, *931*, 1–16.

Wertheimer, M. (1923). Investigations into the lessons of Gestalt psychology. *Psychological Research*, *4*, 301–350.

Westen, D. (1996). *Psychology: Mind, brain, and culture*. New York, NY: Wiley.

Whorf, B. L. (1956). *Language, thought, and reality: Selected writings*. Cambridge: Technology Press of MIT.

Wickens, D. D. (1972). Characteristics of word encoding. In W. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 191–215). New York, NY: Winston.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in my insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*, 655–664.

Wiener, N. (1950). Cybernetics. *Bulletin of the American Academy of Arts and Sciences*, *3*(7), 2–4.

Wiener, N. (1961). *Cybernetics: Control and communication in the animal and the machine*. Cambridge, MA: MIT Press.

Williamson, T. (2008). *The philosophy of philosophy*. New York, NY: Wiley-Blackwell.

Willingham, D. T., & Dunn, E. W. (2003). What neuroimaging and brain localization can do, cannot do, and should not do for social psychology. *Journal of Personality and Social Psychology*, *85*, 662–671.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after 100 ms exposure to a face. *Psychological Science*, *17*, 592–598.

Wilson, E. O. (1999). *Consilience: The unity of knowledge*. New York, NY: Random House.

Wilson, F. A., O'Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, *260*, 1955–1958.

Wilson, S. W. (1985). Knowledge growth in an artificial animal. In Proceedings of the first international conference on genetic algorithms and their applications (pp. 16–23). Hillsdale, NJ: Erlbaum.

Wilson, T. D., Lindsey, S., & Schooler, T. Y. (2000). A model of dual attitudes. *Psychological Review*, *107*, 101–126.

Wiltschko, R. (1996). The function of olfactory input in pigeon orientation: Does it provide navigational information or play another role? *Journal of Experimental Biology*, *199*, 113–119.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences USA*, *104*(19), 7780–7785.

Windholz, G., & Lamal, P. A. (1985). Kohler's insight revisited. *Teaching of Psychology*, *12*, 165–167.

Winfree, A. T. (1984). The prehistory of the Belousov-Zhabotinsky oscillator. *Journal of Chemical Education*, *61*(8), 661.

Winner, E. (2018). *How art works: A psychological exploration*. Oxford, England: Oxford University Press.

Winner, E., Brownell, H., Happé, F., Blum, A., & Pincus, D. (1998). Distinguishing lies from jokes: Theory of mind deficits and discourse interpretation in right hemisphere brain-damaged patients. *Brain and Language*, *62*(1), 89–106.

Witt, J. K., & Proffitt, D. R. (2008). Action-specific influences on distance perception: A role for motor simulation. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1479–1492.

Wittgenstein, L. (2009). *Philosophical investigations*. Hoboken, NJ: Wiley-Blackwell. (Original work published 1953)

Wojciulik, E., & Kanwisher, N. (1999). The generality of parietal involvement in visual attention. *Neuron*, *23*, 747–764.

Wojnowicz, M. T., Ferguson, M. J., Dale, R., & Spivey, M. J. (2009). The self-organization of explicit attitudes. *Psychological Science*, *20*(11), 1428–1435.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, *9*(1), 33–39.

Wood, N. L., & Cowan, J. (1995). The cocktail party phenomenon revisited: Attention and memory in the classic selective listening procedure of Cherry (1953). *Journal of Experimental Psychology: General*, *124*, 243–262.

Wooldridge, M. (2000). *Reasoning about rational agents*. Cambridge: MIT Press.

Wu, L. L., & Barsalou, L. W. (2009). Perceptual simulation in conceptual combination: Evidence from property generation. *Acta Psychologica*, *132*(2), 173–189.

Wynne, C. D. (2001). *Animal cognition: The mental lives of animals*. New York, NY: Palgrave Macmillan.

Yee, E., & Thompson-Schill, S. L. (2016). Putting concepts into context. *Psychonomic Bulletin & Review*, *23*(4), 1015–1027.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology*, *18*, 459–482.

Yoshimi, J. (2012). Active internalism and open dynamical systems. *Philosophical Psychology*, *25*(1), 1–24.

Zadeh, L. A. (1965). Fuzzy sets and systems. In J. Fox (Ed.), *System theory* (pp. 29–39). Brooklyn, NY: Polytechnic Press.

Zangaladze, A., Epstein, C., Grafton, S., & Sathian, K. (1999). Involvement of visual cortex in tactile discrimination of orientation. *Nature*, *401*, 587–590.

Zanna, M. P., & Olson, J. M. (1982). Individual differences in attitudinal relations. In M. P. Zanna, E. T. Higgins, & C. P. Herman (Eds.), *Consistency in social behavior: The Ontario Symposium* (Vol. 2, pp. 75–104). Hillsdale, NJ: Erlbaum.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*, 37–46.

Zhang, S., & Schmitt, B. (1998). Language-dependent classification: The mental representation of classifiers in cognition, memory, and ad evaluations. *Journal of Experimental Psychology: Applied*, *4*(4), 375–385.

Zingg, B., Hintiryan, H., Gou, L., Song, M. Y., Bay, M., Bienkowski, M. S., . . . Dong, H.-W. (2014). Neural networks of the mouse neocortex. *Cell*, *156*(5), 857–859.

INDEX

abstract concepts, 135–36, 178, 275, 319, 444–45, 465–66
abstraction, 8, 356, 364, 413, 420
accuracy, reduced, 441
actions, 6–10, 18, 20–21, 137, 140–42, 181–84, 240–41, 276–77, 311, 324–25, 327, 335–36, 341, 350, 352–53, 408–10, 421, 427, 439, 459–61
  synaptic, 306
action schemas, 184, 188
activating motor cortex, 440–41
activation value, 193–94
actuators, 359, 408, 410–11, 422
ADHD, 174–75, 187
agents, 355, 359–60, 377–78, 380, 383–84, 387–88, 407–11, 421–23, 425, 427, 461, 463–64, 469
AGI (artificial general intelligence), 362, 386–87, 389, 395, 397, 400, 414, 421
algorithmic level, 9, 468–69
algorithms, 8, 10, 53–54, 293, 355, 359, 361, 374–75, 378–79, 407, 411–12, 414–18, 435–36
ALife, 235–37
American Sign Language (ASL), 265, 287–88
American Society for the Prevention of Cruelty to Robots (ASPCR), 435
amodal, 135, 438, 442, 444–45, 454, 457
amounts of attention, 105–6
amygdala, 175, 177, 299, 304–5, 318, 329, 332, 334, 339, 344–45, 349
amygdala activation, 334, 349
analogies, 2–3, 5, 7, 9, 23, 65, 67, 114, 139, 167, 311–12
animal behaviors, 80, 360, 370
animal lesion studies, 148, 187
animals, 4–5, 18–21, 78, 80–81, 147–49, 175, 205–9, 217–19, 229–34, 237–41, 264–65, 286–89, 294–95, 300, 304, 323, 325, 382–84, 398–99, 402–3
animal species, 205, 229, 234–35, 239, 241–42, 261, 411
animal species change, 230–31
animal studies, 179, 404
Animat, 236–37
ANNs (artificial neural networks), 18, 134, 189–93, 197, 200–203, 226, 280, 293, 360–62, 365, 452–53
anterior attention systems, 324–25
apperceptive agnosia, 162–63, 187
apperceptive agnosic patient, 162
apple, 3, 21–22, 116, 192, 194, 268–69, 272–74, 287–88, 327, 386, 438
approaches
  dynamical, 446, 458
  neurobiological, 51, 56
architectures
  parallel processing, 191, 222
  serial processing, 191
areas, parietal, 325, 330, 332
arousal, 50, 105–6, 156, 169, 172–73, 298–99, 304, 308, 315, 336
artificial general intelligence. *See* AGI
artificial neural networks. *See* ANNs
ASL (American Sign Language), 265, 287–88
ASPCR (American Society for the Prevention of Cruelty to Robots), 435
attention, 60–61, 66–67, 83, 85, 96–101, 104–9, 113, 141, 167–74, 183–85, 187–88, 307, 309, 315, 317–18, 323–26, 328–29, 337–38, 465
  concentrated, 98
  disengage, 170, 172
attentional, 168, 325, 329
attentional resources, 60, 104, 124, 169
attentional theories, 107
attention-deficit hyperactivity disorder, 174
attention systems model, 324–25
attitudes, 20, 321, 329, 334–42, 353
attractors, 446, 450–52, 458
attributions, 20, 321, 329, 341–45, 353
autism, 20, 153, 321, 324, 332–33, 437
autobiographical memories, 309, 439
automatic attentional processes, 183–84

backpropagation, 189, 195–97, 199, 293, 361, 382, 385, 411–12, 416
Baddeley, 123–25, 145, 179
basic emotions, 223, 299–301, 317
behavioral economics, 19, 229–30, 253, 262
behavior and experience, 321–22
behaviorism, 16, 35, 62, 78–80, 82–83, 362
  psychological, 28
behaviorist account, 263
behaviorists, 16, 58, 78–81, 83, 471
behaviors, 16–17, 28–30, 39, 57–58, 76–81, 148–49, 235, 237, 240–41, 243, 260, 315, 320–22, 335–37, 341–42, 350–53, 361–62, 408–11, 427, 450–52
  emotional, 20, 297, 314
  organism's, 82, 175–76
  people's, 321, 343
  problem-solving, 136
  swarm, 427
beliefs, 26, 31, 34–35, 240–41, 251, 330, 333–36, 339–41, 350, 353, 367, 371, 427
bioinformatics, 419
birds, 95, 133, 201, 205–9, 229, 232–34, 237–38, 244–45, 282, 368, 407
Black people, 349
Black person, 348–49
blocks, 101, 141–42, 342, 425
Blue Brain Project, 391, 393
body, 21, 26, 28, 30–33, 56, 150, 158–61, 165–66, 172–73, 387, 411, 421, 431–32, 436–40, 457–58
  cell, 154–56, 158, 169, 171, 187
body parts, 153, 243, 392, 467
body problem, 14, 25–26, 28, 45, 56, 361
body size, 237–38, 307, 323
bottleneck theories, 101
brain, 17–23, 26–36, 44–49, 51–52, 55–56, 147–51, 159–62, 166–69, 175–76, 185–88, 211–13, 219, 237–40, 291–93, 360–62, 389–92, 399–403, 431–34, 436–38, 452–58
  child's, 277
  person's, 148, 466
  real-world, 189, 226
brain action, 52, 187
brain activity, 17, 28–29, 52, 59, 74, 148–51, 202, 327, 432, 456, 465
  record of, 178–79
brain anatomy, 17, 156–57
brain areas, 148, 150, 152, 172, 175, 177, 179, 181–82, 281–82, 290–92, 325, 327–30, 332, 344
  distinct, 170, 179
  multiple, 170, 173
brain damage, 19, 148, 172–73, 182, 187, 263, 284, 289, 296
brain-damaged patients, 18–19, 59, 85–86, 292
Brain Emulation, 392
brain function, 144, 147, 152, 394
brain imaging studies, 188, 330
brain imaging techniques, 292
brain mechanisms, 149, 260
brain regions, 10–11, 149, 151–53, 158, 162, 170, 172, 175, 179, 181, 389, 391, 456–57
  motor, 446, 457
brain size, 237–39
brain states, 29, 35, 457
  physical, 29, 34
brain structures, 150, 171, 175, 186–87, 238, 240, 245, 318, 339, 349
brain tissue, 148, 152, 158, 184, 187
brainwaves, 403, 432
Broadbent's Filter Model, 101, 103
Broadbent's model, 101–3
Broca's area, 179, 289, 291–92, 413

Cambridge, 56, 109, 145, 227, 262, 296, 318, 400, 473
Cannibal problem, 139, 141
capabilities, 5, 123, 366, 370, 387–89, 396, 400, 403, 409, 421, 423
capacity model of attention, 105, 107, 170
capacity models, 101, 105, 107, 109
capacity theories, 101, 107
categories, 3, 5, 8, 132–34, 148, 205, 207–8, 244, 255, 323–24, 338, 341, 347, 351, 353
causal relationship, 26, 38
cells, 149, 153, 155–56, 165–70, 176, 187, 189, 194, 239, 325, 327, 389, 391, 394
chaotic systems, 447–48
characteristics of semantic networks, 203
children, 61, 174–75, 246, 249, 252, 255–56, 278–79, 282–83, 319, 323–24, 330, 334, 415–16
chimps, 73, 234–35, 238, 287–88, 295
Churchland, 10, 29, 50–51, 56
classical cognitive science, 23, 437, 445–46, 451, 454
classical theory of concepts, 132
cluster, 39, 181, 220–21, 378

cognition, 20–22, 85, 229–30, 232, 237–38, 297–300, 306, 317, 319–22, 352, 437–38, 445, 457–58, 460–61, 463–67, 469–70, 473
cognition interaction, 306–7
cognitive abilities, 41, 149, 229, 261–62, 406
cognitive approach, 16, 81, 83–109, 136, 144, 147, 182, 186, 321, 471
cognitive computations, important, 455
cognitive demons, 89–91
cognitive dissonance, 319, 336–38, 342, 353
cognitive economy, 208
cognitive neuroscience, 17, 132, 147, 186, 188, 321
cognitive operations, 136, 147–48, 150, 181, 438, 461, 466, 473
cognitive perspective, 83, 85, 100, 321
cognitive processes, 17, 144, 152, 186, 243–44, 261–62, 304, 306, 317, 319, 326, 328, 466–67, 469, 472
  human, 361, 365
cognitive psychologists, 16–17, 22, 55, 57–58, 84, 203, 245, 366, 398, 467–69
cognitive psychology, 13, 16–19, 22, 80, 83–84, 111, 117, 123, 144, 258, 264
cognitive skills, 229, 234–35
cognitive systems, 55, 141, 315, 445, 454, 457, 459
cognitive theories, 85, 111, 471
color, 36–37, 70–72, 94, 96, 98–99, 161–62, 167, 169–70, 179, 182, 307–8
color perception, 37, 85, 285–86
commands, 118, 127, 160, 276–77, 288, 291, 294, 369–70, 424, 431, 446
complex objects, 93, 214
component process model, 170, 172, 187
computational elements, 355, 359, 375, 400–401
computational modeling, 17
computations, 2, 7–10, 12, 16, 23, 35–36, 53, 153, 190–91, 451, 454–55, 457, 459–60
computer capacities, 372, 387
computer models, 19, 84, 318, 347, 361, 374
computer mouse, 244, 463
computers, 2–3, 9–10, 16, 20, 22–23, 34–36, 84, 94, 312, 372–75, 388–90, 392, 394–97, 400, 407–8, 420–21, 431–32, 434–35
computer technology, 386, 389–90
conceptions, 10, 44, 46, 49, 256, 281
concept representation, 5, 135
concepts, 3, 5–6, 12, 18, 36–37, 55–58, 111, 132–36, 192, 203, 207–10, 223, 244–45, 275–77, 285, 364, 366, 438–39, 442–45, 471
  concrete, 135–36, 444
conceptual metaphors, 136, 275
conceptual representation, 111, 132, 134–36, 144, 166, 281, 438
conditioned stimulus, 79
connectionist approach, 12, 200–201
connectionists, 10, 12, 22–23, 58, 85, 189–90, 192–93, 245, 355, 467
conscious experience, 19, 28, 36–37, 42, 44, 47, 49–52, 63, 65, 74, 430
consciousness, 14–15, 25–26, 41–54, 56, 65, 67, 76, 79–80, 212, 360–61, 364, 387, 390, 403, 405
  human, 362
consolidation, 121, 176–77, 181, 188
contention scheduling, 184–85
context, visual, 268, 272, 274
continuity of mind, 454, 457
control, cognitive, 152, 181, 184, 222, 292
controlled attentional processes, 183, 343
control of attention, 170–71, 326
correct solution, 73, 137, 145, 247
cortex, 49–51, 147, 153, 157–59, 161, 165, 167, 169, 175, 177–78, 180–81, 290, 304–5
cortical areas, 49–50, 219, 291, 441
creatures, 12, 30, 235–37, 405
cultures, 77, 260, 299, 317, 323, 353, 398, 469–71
curvature, 28, 72, 96, 98, 349

damage, 85–86, 148–49, 162, 165, 172–73, 177, 181, 185–88, 200–201, 219, 226, 289–92, 310
declarative knowledge, 7, 119, 145
Declarative memories, 181, 188
declarative memory, 119, 181
deep learning, 293, 373, 375–76, 382–83, 411–12, 436
Deep learning. *See* DL
defects, 303, 346–47, 413, 418
defining features, 111, 132–33, 448
  list of, 132
deliberative architectures, 401, 423–25, 427
dendrites, 154–56, 306, 356
Dennett, 46–48, 53, 56, 212
Descartes, 30–32, 37, 46
design, 16, 84, 104, 364, 368–70, 397, 404, 418, 420, 435, 437–38
diagnosis, 175, 368, 378–81, 435

disciplines, 1–2, 12–14, 16, 19, 22, 25, 55, 57–58, 78, 81, 84, 144, 147, 357, 467–73
disease, 1, 190, 220–21, 380, 387, 396
disgust, 40, 300, 302, 315, 317, 327–28, 351
disorders, 86, 153, 156, 161–63, 165, 173–74, 176, 182, 302–3, 330, 333–34
disorganized thinking, 220–21, 227
distraction, 60, 101, 301
distractors, 96–99, 108, 167
distributed network models, 170, 172, 187
divergent thinking, 220–22, 227
DL (Deep learning), 293, 365, 373, 375–77, 382–83, 385–86, 388, 390, 411–13, 415–16, 435–36
DL systems, 383–86, 412–13
doctors, 204–5, 289, 340, 464
domain-general mechanisms, 279–80, 282
dualism, 25, 27, 30, 32–34, 56
dual-process model, 340
dynamical systems approach, 12, 446, 453, 472
dynamical systems perspective, 451

EA (Evolutionary algorithm(s)), 414–19, 435–36
EEG, 17, 147, 149–51, 187, 433, 465
electrical stimulation, 152, 158
elephant, 12–13, 237–39, 269
embodied cognition, 22, 275, 437–40, 445–46, 457, 472
embodied cognitive science, 403, 405, 436
emotional approach, 13, 297–317
emotional states, 223, 306, 313, 315, 327, 334, 341
emotion and memory, 304, 308
emotion nodes, 224
emotions, 19, 22, 223–26, 297–302, 304, 306–15, 317–18, 329, 335, 350, 353, 430, 432
  human, 312, 396, 430
encoding, 121–22, 127, 145, 305, 309–10, 393
engineers, 43, 56, 61, 249–50
England, 56, 109, 145, 227, 262, 264, 318, 353
English, 55, 198, 264–65, 267, 269, 275, 282, 405
environment, 14–15, 20–21, 35, 39, 58, 107–8, 229, 233–34, 236–37, 241–42, 245–46, 260–61, 320, 369–70, 403–11, 420–21, 424–25, 427, 452, 459–66
Environmental signals, 359, 451
EP (Evolutionary psychology), 19, 40, 68, 229–30, 241–43, 245, 255, 258, 260–62
episodic memory, 119, 123, 181, 188
ERPs (event-related potential), 104, 149, 187, 339
error signal, 194, 196–97, 200, 202
event-related potential. *See* ERPs
Evolutionary algorithm(s). *See* EA
Evolutionary psychology. *See* EP
evolved psychological mechanisms, 229–30, 242–43, 249, 251, 258, 260–62
executive function, 170, 180–81, 183–85, 330, 352
experience, 36–39, 41–44, 46–47, 49, 52–53, 56, 62–67, 178, 223, 321–22, 327, 374, 408, 410, 444
experimental group, 59–60
experimental philosophy, 15, 22
experimental results, 19, 30, 203, 469–70
expert systems, 5, 361, 364, 368, 371, 424
expressions, 4–5, 23, 166, 270, 300
extended cognition, 437, 459–60, 464–67
eyes, people's, 465

fear, 28, 30, 78, 298–300, 302, 304–6, 313, 315, 317, 334, 350
feature demons, 89–91
feature detection models, 88, 90, 107
feature extraction, 96, 107
feature integration, 96, 108
features, 18, 32, 45, 51, 88–90, 92–94, 96, 98–100, 107–8, 132–35, 160–65, 167, 203, 213, 373–74, 438
  complex, 90, 92
  simple, 89, 373
FFA (fusiform face area), 52, 166
field theory, 58, 74–75
filter, 101–3, 169, 307
folk psychology, 29, 56
food items, 179, 236, 287–88
food object, 462
formal logical systems, 4–5
friend, 12, 176, 181, 335, 346–47, 350, 465
functionalism, 16, 25–26, 34–36, 43, 58, 62, 66–69, 362
functionalists, 53, 56, 58, 66, 69, 230, 437
fusiform face area. *See* FFA

General Problem Solver. *See* GPS
General Problem Solver Model, 139–40
Geschwind model, 289, 291–92, 295
gestalt psychologists, 58, 64, 66, 69, 71, 73–75, 82, 136
goals, 76, 136–42, 145, 147, 183–85, 310, 312, 314, 344–45, 365–66, 375–77, 386–87, 409, 411, 424
  desired, 141, 241
Goertzel, 386–89, 435
GPS (General Problem Solver), 139–41, 145, 406–7

GPUs (graphical processor units), 365
grammar, 200, 265, 271, 277–78, 287, 295
graphical processor units (GPUs), 365
grouping, 71–72, 75, 116, 142, 162, 244, 379
group of people, 119, 212, 349
groups, 59–60, 71–72, 75, 116, 153, 166–67, 182, 212–13, 230, 253, 275, 285–87, 333–34, 338, 344–45, 348–51, 378–80

hand-action verbs, 441–42
Hanoi problem, 138–39, 141, 182
happiness, 19, 28, 223–24, 298, 300, 303–4, 313, 317
hard problem of consciousness, 42
hardware, 9, 17, 147, 360, 363–64, 394, 408–9
heuristics, 138, 140–41, 145, 225, 249, 253, 262, 371, 417
hidden layer, 195–96, 198–99, 208
hierarchical networks, 212–13
hierarchical semantic network model, 207
hierarchies, 205, 208, 213, 271, 405
hippocampal system, 176, 180–81
hippocampus, 41, 147, 166, 175–78, 181, 187–88, 221, 305
human beings, 84, 140, 166, 244, 251–52, 282, 361, 396–97, 400, 402, 436
human brain, 152, 156–57, 200, 202, 208, 359, 362, 386–87, 390–91, 395, 399–400, 410–11, 431–33
human cognitive abilities, 256, 360, 366
human cognitive behavior, 401, 411, 436, 456
human intelligence, 322, 358, 388, 400, 408, 414, 435
human interactions, 313, 321, 401, 428, 431
human mind, 1–2, 17, 29, 84, 208, 212, 229, 241–42, 355, 359, 361
humanoids, 404, 407, 412, 414, 423
human participants, 18, 294, 327
humans, modern, 230, 251
human-tool system, 463
hundred people, 254

iconic memory, 112–14, 144
idealized results, 114–15, 129
IFR (International Federation of Robotics), 405
IJA (initiating joint attention), 324, 326
illness, 367, 380–81, 396, 402, 436
image, 52, 87–89, 94, 96–97, 107–8, 123, 125–29, 131, 151–52, 165–66, 274, 365, 442, 444–45, 457–59
image processes, 127
imagery, 17, 85, 111, 125–26, 132, 150, 274, 285
image schemas, 263, 274–75, 442
images of objects, 126
image structures, 126–27
image transformations, 128, 131
implementational level, 9–10, 17, 147
impressions, 20, 321, 328, 340–41, 353
inability, 121, 162–63, 165, 177, 188, 195, 291, 358
independent variable, 57, 59–60, 81, 113
infants, 39, 256, 279–81, 302, 314, 323–24, 405
information, 2, 21–23, 46–52, 84–87, 100–104, 107–9, 111–17, 120–27, 143–45, 158–61, 166–69, 176–77, 186–87, 189–90, 245–46, 279–82, 337–40, 347, 364–67, 459–61
  consistent, 338
  contextual, 243, 294, 342
  new, 12, 41, 44, 176–77, 188, 192, 225–26, 247, 293, 305, 310
  processed, 172, 451
  raw, 384
  visual, 123–24, 158, 160, 178, 187, 280, 286, 291, 459
informational sources, 100–101
information changes, retinal, 22
information flow, 145, 217, 291, 317, 454, 469, 473
information in long-term memory, 87, 121, 127, 203
information in memory, 121, 226, 339
information in working memory, 115, 117
information processing, 2, 10, 16, 23, 78, 84–85, 121, 144, 160, 360–61, 364
  human, 84–85
information-processing models, 22, 84, 102, 147, 453
information-processing systems, 7, 111, 361
information processors, 2, 9, 22, 84, 470
information representation, 112, 356
information sources, 100–101, 109
ingroup, 348–49, 353
Inhibitory networks, 210
initiating joint attention. *See* IJA
input layer, 195–96, 198–99, 209
input units, 196–98
insight learning, 73, 75, 136, 234
intelligence, 61–62, 237, 239–41, 252–53, 322–23, 356, 359, 362–63, 365–66, 374, 400–401, 405–9, 411, 414, 420
Intelligence Quotient (IQ), 61, 388
intelligence tests, 61
intelligent agents, 355, 359–60, 362, 366, 368, 372, 388–89, 400–401, 406, 408–11, 414, 434, 436
interaction-dominant dynamics, 189, 211, 455–56, 463

interactions, 13–14, 16, 37, 39, 121, 143, 145, 291–92, 314–15, 406, 408, 427–28, 454–57, 461–62, 467
  causal, 45
interdisciplinary approach, 1–2, 23, 321, 352
International Federation of Robotics (IFR), 405
introspection, 34, 44, 46, 57–58, 62, 64–66, 69, 79–81
IQ (Intelligence Quotient), 61, 388
items, 98–99, 113–19, 121–22, 165, 169, 173, 210, 244–46, 250, 307, 309, 369, 388
iterations, 374, 384–86, 418

joint attention, 319, 323–26, 352

Kahneman's capacity model of attention, 105–6
Kismet, 312, 314–16, 430
knowledge, 7, 10, 14, 25–26, 37–40, 56, 137, 141–42, 190, 204–5, 363–64, 366, 398–99, 408, 414, 444, 452, 467–68
knowledge of concepts, 132
knowledge representation, 18, 84, 189, 208, 237, 370
Kosslyn, 111, 126–28, 132, 145

language, 1–2, 8, 10, 17, 19, 55, 57, 201, 252–53, 262–66, 269–70, 274–75, 277–88, 292–96, 441–42
  human, 270, 276, 281
  second, 282
language ability, 19, 86, 160, 263–64, 282, 289, 296
language acquisition, 16, 19, 263, 277, 279, 282, 284, 295, 324
language and thought, 19, 263, 295
language capacity, 252
language development, 253, 282, 284
language information, 263, 284
language learning, 263, 278, 319, 453
language processes, 442
language processing, 252, 265, 284, 291–92, 296, 440–41, 456
language system, 271–72, 286, 292–93, 330
language task, 125, 443, 445
language use, 19, 136, 276, 286, 295, 324
layers, 51, 87, 196–99, 202, 226, 280, 383–84, 415
learning, 69, 75, 80, 142–43, 147, 175–76, 180–81, 192, 194, 200, 226–27, 260, 277–79, 281–82, 360, 373–77, 386, 400–401, 436–37, 444
learning algorithms, associated, 412
learning process, 260, 325, 374, 414, 436
left hemisphere, 125, 127, 157–58, 160, 179–80, 281, 286, 290, 330, 332, 339
lesion studies, 148, 186
letter display, 113
letters, 87, 89–91, 112–16, 178–79, 192, 196–99, 213, 215, 266, 269–70, 382, 386
levels, 5, 8–10, 17, 28, 52, 54, 205–8, 212–13, 271, 452–54, 466, 469–70, 473
levels of description, 453, 467–68
limbic cortex, 172, 181, 188
linguistics, cognitive, 263, 274–75
links, 7, 136, 193, 196, 200–201, 203, 205, 210, 213, 215, 217, 224–27
listeners, 268, 270, 272, 274, 276–77, 312, 340, 466
listening, 59–60, 265, 278, 282, 291, 442–43, 466
list of parts, 127
loads, cognitive, 336, 339
localist coding, 166, 192
logical operations, 194, 364, 367, 371
logic problems, 249
London problems, 182, 185
long-term memory, 111, 118, 121–23, 127, 141, 145, 176, 180–81, 188, 190, 203, 205

machine learning, 355, 373–77, 383–84, 390, 417, 419
machine learning systems, 377, 399
machines, 36, 53–54, 77, 296–97, 313–14, 356–57, 359, 361–65, 367–70, 372–77, 386–88, 395–96, 398, 401–3, 405–6, 408, 413, 417–20, 431–36
magnetic resonance imaging. *See* MRI
malfunctioning, 463
malfunctions, 370, 430
mammals, 4, 207, 237–38, 323
meaning, 3, 5, 8, 54, 84–85, 116, 126, 135, 203, 265, 268–70, 274–75, 287–90, 294–96, 444
mechanistic explanations, 452–53
members, 5, 132–33, 231, 234, 236, 253–54, 287, 289, 348, 350–51, 353, 415, 427
  ingroup, 348–49
memorize, 59, 117, 128, 247, 328, 441
memory, 16–20, 41, 55, 57–58, 60, 87, 102, 111–13, 118–19, 121–23, 141–45, 147, 175–76, 187–88, 233–34, 243–46, 308–9, 317–18, 337–39, 364–65
memory abilities, 60, 176–77
memory deficit, 176
MEMORY MODELS, 112, 121
memory performance, 59–60
memory processes, 60, 175, 308
memory search, 117, 119, 210
memory skills, 41

memory strategies, 41
memory system, 112, 118, 125, 208, 308
  distinct, 112, 143
memory task, additional verbal, 286
memory types, 118, 121, 143, 145
mental activity, 16, 47–48, 67, 100, 437, 454, 460–61, 466, 472
mental operations, 8, 22, 58, 67, 189
mental processes, 7–8, 17, 58, 62, 66–67, 83, 85, 117, 237, 280, 458
  human, 19, 190
mental representations, 1–3, 5, 7, 74, 80, 125, 269, 275, 286, 459, 461, 467
mental states, 28–29, 32, 35–36, 41–42, 44, 46, 56, 312–13, 328, 330, 333, 344
messages, 101, 153–55, 187, 216–17, 220
mind, 1–5, 12–14, 16, 18–23, 25–37, 41–42, 44–47, 54–58, 61–67, 75–85, 133–34, 144, 221–22, 241–43, 296–98, 328–30, 359–62, 454–58, 466–69, 471–73
mind and behavior, 16, 57
mind and body, 26, 31–32, 56, 405, 436
mirror neurons, 20, 325–26, 352
mirror neuron system, 319, 325, 327–28, 439
modal memory model, 111, 121–22
modal model, 121–23, 141, 145
modal representations, 135, 457
models, 77, 79, 87, 89–90, 102–5, 107–8, 123, 139, 141–45, 170, 172, 184–86, 208–10, 222–23, 291–92, 313–14, 373–75, 383–84, 389–91, 469–70
  connectionist, 192, 200–201
  mathematical, 382–83, 453
models development, 362
models of pattern recognition, 107, 160
modularity, 16, 86, 282, 454–55
modularity of mind, 83, 85, 109
modules, 16, 19, 85–86, 104, 108, 111, 243, 270, 280, 282, 423, 455–56, 458
molecules, neurotransmitter, 156, 187
monkeys, 113, 165, 169, 172, 179, 181, 238, 241, 287, 325, 327, 332, 462–63
mood-congruent memory, 309
mood-dependent memory, 309–10
motor actions, 173, 457, 459
motor cortex, 135, 153, 161, 181, 439, 441
  hand region of, 441
movements
  cognitive, 80, 83
  functionalist, 68–69, 82
MRI (magnetic resonance imaging), 389–90, 432
multimode model, 101, 104, 108–9
multimode model of attention, 104
multiple types of intelligence, 241
muscles, 40, 233, 291, 407, 432–34
mutations, 41, 415–17, 419

natural language processing (NLP), 293–94, 374, 419
natural language understanding, 388, 413
natural selection, 18–19, 67, 229–32, 258, 261, 414–15
nature of language, 263–64
negative evidence, 277–78
negative words, 310, 348–49
neocortex, 238, 252, 323, 370, 420
nervous systems, 10–11, 157, 175–76, 357–58, 392, 395, 400, 451, 462, 464, 466
NETtalk, 197–99
network models, 200, 223, 226, 263
networks, 10, 18, 51, 121, 123, 144, 169–70, 189–90, 192–205, 208–12, 215–17, 219–20, 222–23, 226–27, 446, 451–53
  artificial, 193, 200, 202, 245
  biological, 194, 200, 202
network science, 18, 189–90, 211–12, 222, 226, 446
network structure, 211
neural models of attention, 170
neural networks (NNs), 10–11, 184, 189–90, 192–94, 201, 208, 220, 373–74, 377, 382–83, 385, 390, 424–25
neural processes, 35, 328, 464
neural representations, 107, 135, 173, 440–41
neurobiological approach to consciousness, 51, 56
Neuron model, 375, 419
neurons, 9–11, 49, 90, 147–49, 152–55, 158, 165–68, 189–90, 192–94, 200–202, 212–13, 218–22, 240, 305–6, 325–27, 382–84, 389, 391–92, 394–95, 414–15
  active, 17, 151
  cortical, 34, 49
  individual, 10, 17, 45, 148–49, 166, 186
  multisensory, 462
neuroscience, 17–20, 22, 49, 51, 58–59, 144, 147–48, 181, 183, 186, 261, 263–64, 330, 469, 471
Neuroscience Approach, 17, 147–87, 263, 329
neuroscience of memory, 175
neuroscience perspectives, 147, 320
neuroscientists, 17, 43–44, 147, 321, 377, 391, 453, 467–69

neurotransmitters, 17, 155–56, 306
  major, 156
New York, 56, 82, 109, 188, 227, 262, 296, 318, 353, 400, 473
New York Times, 355–56, 387, 402, 433–34
next input, 45, 199–200, 446, 448
NLP. *See* natural language processing
NNs. *See* neural networks
nodes, 18, 138, 192–96, 198–203, 205, 208–12, 215–19, 222–24, 226–27, 373, 375, 417–18, 452–53
  hidden, 209, 293
  single, 192, 226
nonhuman animals, 232, 234, 241, 286, 288
nonlinear activation functions, 194, 200
Norman Memory Selection Model, 101, 103, 108
noun phrase. *See* NP
NP (noun phrase), 271–72, 274
number
  acceptable, 380, 385
  equal, 59–60
  phone, 114–15, 194
  single, 61

objectify, 62, 81
objective account, 42, 44
object parts, 94, 96, 108, 168, 384
object permanence, 233
object recognition, 22, 96, 98, 160, 179, 187
  visual, 147, 158, 162, 439, 456
object representations, 92, 95, 98, 162, 165
objects, 5, 36–37, 65, 86–88, 90–94, 96, 125–26, 128–29, 160–69, 172–73, 179–80, 256–57, 287–88, 320, 324–25, 332–34, 383–84, 421–24, 442–43, 461–62
  inanimate, 319, 330
  physical, 32, 125, 334
  second, 128
  visual, 82, 162, 192, 439–40
objects and events, 457, 465
operant conditioning, 80, 385, 425, 433–34
operators, 7, 136, 138–39, 141–43
ordered networks, 215–16, 219
organic brain, 53, 392, 397
organism, 7–9, 43–44, 67, 69, 78–79, 242–43, 258, 260, 298, 300, 400, 403, 452, 460, 462
outgroups, 348, 350, 353
output layer, 195–96, 198–200, 209, 226
outputs, 2, 4, 84, 168, 190–91, 193, 197–98, 243, 291, 394–95, 406, 408, 446, 448
  net, 193–94
output units, 196–97
Oxford University Press, 56, 109, 145, 188, 262, 318, 353

pandemonium model of pattern recognition, 90–91
parameters, 17, 84, 132, 371–72, 376, 380, 384–86, 412, 414, 432, 436
parietal lobe, 158, 161, 167, 169–70, 172, 187, 292
participants, 15, 17, 59, 96–98, 101, 112–17, 127–32, 134, 149–51, 169–70, 178–79, 204–7, 249–51, 254–56, 299–300, 307–9, 328–30, 336–40, 438–44, 465–66
  white, 348–50
Particle Robots, 428–29
parts, 6, 8–10, 16, 18–22, 32–34, 44–45, 48–50, 54–56, 69–72, 74–76, 92–94, 125–28, 148–50, 160–62, 165–70, 175–76, 212–13, 271–72, 289–92, 347–49
  object's, 96, 127
patients, 85–86, 148, 160–65, 172–74, 176, 181–83, 185, 187, 289–90, 292, 296, 310–11, 332–34, 379–81, 464
patriotic state of mind, 35
pattern of activation, 166, 189, 192, 196, 198, 202, 226
pattern recognition, 83, 86–87, 90–91, 93–94, 96, 102, 107–8, 149, 158, 160, 167
pattern recognition process, 87–88
patterns, 10, 21–22, 51–52, 86, 88, 102, 116–17, 128, 130–31, 168, 201–2, 293, 374, 386–87, 428
patterns of brain activity, 52, 148, 151, 327
people, 5–6, 20–23, 133–36, 215–17, 229–30, 247–51, 253–55, 302–3, 308–10, 315–17, 319–21, 323–24, 327–29, 333–38, 340–42, 344–53, 380–81, 444–45, 463–66, 469–71
  white, 349
people's states, 20, 332
people's states of mind, 20, 332
perceptrons, multilayer, 195–96, 383
perceptual organization, 70–71, 75
perceptual simulations, 134–36, 438, 442–44
person, white, 348–49
perspectives, memory's, 49
PET scans, 150
phenomenal concept of mind, 42
philosophers of mind, 14, 466
philosophical behaviorism, 28

philosophy, 13–15, 22, 25, 47, 51, 55–56, 247, 250, 263–64, 357, 361
philosophy of mind, 26, 35, 56
phonemes, 265–70, 295
physical processes, 9–10, 30, 53
physical states, 27, 29, 35, 56
physiological changes, 298, 317
positive words, 348–50
posterior parietal cortex, 172–73, 179, 344–45
preattentive, 97
preferences, 40, 142, 281, 333, 341
prefrontal cortex, 134–35, 179, 184–86, 310, 393, 413
prejudices, 20, 319, 321, 329, 350–51, 353, 371, 431
premises, 6, 38, 247, 368, 400
primary motor cortex, 158, 181, 291
primates, 237–38, 252, 287–89, 322–23, 418, 462
principles, evolutionary, 230, 236
prism model of social cognitive neuroscience, 319, 322
probability, 6, 72, 217, 236, 246, 249, 251, 254, 366, 380–81, 385
  occupation, 220
problems, 8–10, 19, 44–47, 73–74, 136–43, 182–86, 191–92, 201–2, 229–30, 234, 241–43, 247, 249, 260–62, 364–65, 388, 407–9, 414–15, 417–18, 464–65
  block-world, 141–43
  bouncer, 248
problem solving, human, 139, 356
problem space, 111, 137–38, 141–42
procedural memory, 118, 181
processes
  basic, 471
  motor, 438, 442, 445–46, 472
  preattentive, 97
process information, 100, 175, 394
processing, 9–10, 47–49, 83–86, 93–94, 100–104, 108–9, 121, 123–24, 169, 189–90, 225, 279–81, 313–14, 441, 455–56
  attentional, 187, 304
  emotional, 311, 313
  guide, 225–26
  open, 225
  underlying attribution, 344–45
processing centers, 455
processing units, 89, 191, 222
process models, 16, 84–85, 121, 311, 453, 455, 458, 469
  cognitive, 85, 469
production rules, 6, 121, 139, 141, 143, 184, 236, 364, 452
production systems, 6
properties, 32, 45–46, 51, 201, 205, 207–8, 243, 245, 378, 387, 408–9
properties of concepts, 443
propositions, 3, 5–7, 23, 32, 37, 223, 285, 359
proximity, 70–73, 75, 205, 209
Psychoanalytic psychology, 16, 75–76, 78
psychological concept of mind, 42
psychological mechanisms, 242–43, 251, 258
psychologists, evolutionary, 19, 41, 68, 230, 242, 258, 260, 470
psychology, 13, 15–16, 22, 57–59, 61–63, 65–69, 74, 79, 81–84, 220, 260, 262–63, 471–72

qualities, 66, 194, 276, 320, 371–72, 405–6

Random networks, 215–16
RAS, 169, 187
reaction time experiments, 134, 269, 274
reaction times, 98, 117–19, 128–29, 131, 205, 328, 348, 441, 453
reagents, 65
reasoning and problem solving, 145, 180
recognition, 86–88, 90–91, 94–96, 102–3, 107–8, 268–69, 366, 370, 377, 411, 420
reconstructive memory, 211
recurrent networks, 51, 451
reduced accuracy for memory, 441
referents, 3–4, 23, 125–26, 288–89, 324, 451, 459, 472
rehearsal, 115, 121–25, 145, 178–80, 188, 194
relevance, personal, 103, 225
representations, 3–10, 12, 23, 89, 93–96, 106–7, 112, 117, 125–27, 135–37, 166, 177–78, 189–90, 192, 195, 202–3, 451–54, 458–61, 467–68, 472
  cognitive, 223
  diagrammatic, 198–99
  motor, 135, 444–45
responses, shared, 466
response times, 129, 131, 172, 205–8
RJA, 324, 326
robotics approach, 21
robots, 5, 20–22, 293, 314–16, 318, 355–56, 361, 392, 396–98, 401–10, 418, 420–21, 423–25, 427–31, 435–36
  individual, 428
rules, 3–7, 10, 45, 53–54, 192, 194, 236, 247–48, 264, 270, 279–80, 364, 370, 427, 451–53

SAS, 184–85, 188
scenarios, real-world people, 344
schemas, 184–85, 225, 310, 347, 350, 353, 471
scientific study of mind and behavior, 57, 81
screen, 9, 46–47, 51, 113, 348–49, 365, 442–44
search, 25, 87, 98, 117–18, 138, 178, 210, 226, 233–34, 293, 417
search function, 97–100, 118
search times, 117–18
Searle, 45–46, 53–54, 56, 276, 400
seizures, 176, 219–21, 227
selection, 101–5, 108–9, 112, 143, 170, 231, 236, 260, 386, 415–16
selection mechanism, 102, 104
selective attention, 100–102, 338
semantic memory, 119, 123, 181, 188, 206–7, 309
semantic network models, 203, 208
semantic networks, 18, 57, 190, 203–5, 208, 210, 226
sensations, 8, 39, 41, 64, 66, 356, 473
sensory information, incoming, 49, 101, 112, 169
sensory memory, 101, 112, 114, 121–23, 144
sentences, 3, 205–8, 263, 265, 267–68, 270–72, 274–79, 287–89, 293, 295, 346, 412, 442–44, 465
sequence, problem-solving, 137
sex differences, 229–30, 256
short-term memory, 51, 102, 104, 114–16, 125, 366
side, 28, 30, 141, 159–60, 170, 172–73, 247–48, 443, 450, 453, 458
  right, 87, 159–60, 173–74, 443
sketch, 5, 94–96, 108, 111, 158
small-world networks, 189, 213, 215, 217–19, 227
small-world networks and synchrony, 219
SOAR, 7, 141–43, 145, 183
SOAR Model, 141–43
social cognition, 20, 319–21, 323–24, 327–29, 352–53
social cognitive neuroscience, 22, 319–22, 328–29, 352
social environments, 242, 249, 322–23
social interaction, 1, 166, 175, 312–13, 323, 328, 332–34, 352
solution, 73, 136–39, 141–42, 167, 182–83, 242–43, 248, 371, 373, 383, 404–5, 414–15, 461, 464–65, 468
solution tree, 137–38
spatial ability, 158, 256
spatial information, 179
speakers, 267, 276–77, 279, 407, 465
species, 30, 34, 40–41, 230, 232, 234–38, 240–42, 244–45, 252, 282, 289, 295
split-brain patients, 41, 160
spoken word recognition, 268–69, 271, 295
spreading activation, 203–6, 209–12
states, 32, 35, 41–42, 137–38, 141–43, 302, 315, 342, 344, 369–70, 373, 394, 424, 427, 449–50
  initial, 137, 140–43, 282
  stable, 202, 454–55
state space, 437, 446, 449–51, 454, 458, 468
statistical learning, 281
statistical patterns, 136, 233–34, 246, 274
statistical signature, 456
steps, 6–10, 60, 84, 86–87, 96, 98, 136–37, 140–41, 196–97, 210, 213, 215–16, 236, 248, 415–16
stereotype content model, 350–51
stereotype content model of prejudice, 350–51
stereotypes, 20, 250, 320–21, 329, 347–48, 350–51, 353
stimulation, 51, 152, 176, 282, 315, 392
stimuli, 62–63, 65–66, 72–75, 79–80, 87–92, 100–105, 108, 112–13, 130–31, 165–66, 195–97, 201–2, 204, 226, 299–300, 302, 309, 339, 348, 425
  fearful, 304–5
  threatening, 299, 304, 338–39
stories, 12, 33, 328–29, 344, 431, 465–66
streams, 47–49, 67, 160–61
structuralism, 57–58, 62, 65–66, 69
students, 62, 65, 69, 100, 201, 335, 337, 340, 384–85, 388, 395
study of mind and behavior, 81, 258
subgoals, 136–39, 183
subjective experiences, 29, 34, 36, 42–43, 51, 53, 56, 62, 69, 317, 327
substances, 26, 28, 31–32, 45–46, 56, 65, 302
subsystems, 16, 270, 295, 455
subtractions, 48, 178–79
superintelligence, 359–60, 397–99
superior colliculus, 169–72, 187
symbolic representations, 4, 135, 191–92, 285, 438, 445
symbols, 3–5, 8, 10, 12, 23, 53–54, 189–90, 192, 264, 288–89, 295, 451–54, 457
symptoms, 143, 149, 173–74, 185, 187, 221, 292, 302, 319, 333, 378–81
  autistic, 333
systems, 8–10, 45, 123–24, 180–81, 211–12, 219, 222, 324–25, 344–45, 362–67, 369–71, 374–77, 385, 390–92, 397–99, 412–13, 446, 448–56, 458, 465–67
  formal, 10, 23
  metacognitive, 186
  physical, 26, 34–35
  slave, 124
systems theory, perceptual symbol, 438

table, 62, 266–68, 327–28, 332, 339, 344–46, 361–62, 367–73, 375–76, 382–84, 386–88, 394–99, 403–9, 412–13, 415, 419–20, 425, 427–28, 451, 454
Tanenhaus, 265, 269–70, 272–74
teacher, 28, 194, 196–200, 340
templates, 83, 87–90, 108, 236
thalamus, 49–51, 169–72, 187, 304–5, 457–58
theories, 26, 34, 37, 57–62, 81, 88, 92–94, 96, 100–101, 107, 121, 126, 131–32, 144–45, 183, 294–95, 332–33, 390–91, 468–69, 471
theories of executive function, 183
theories of mind, 21, 29, 361
theory of mind. *See* ToM
time, retrieval, 206–7
time scales, longer, 265
time series, 456, 463
TMS (transcranial magnetic stimulation), 147, 152–53, 440–41, 456
ToM (theory of mind), 20, 235, 241, 319, 321, 323, 329–30, 332–33, 352–53
tools, 10, 21–22, 114, 147, 149, 260, 263, 270, 274, 431–32, 463–64, 466–67, 472–73
Tower of Hanoi problem, 138–39, 141, 182
Tower of London problems, 182
traits, 18, 37, 41, 78, 231–32, 261, 320, 340, 348, 404
trajectories, 129, 446, 448–52, 454, 458
transcranial magnetic stimulation. *See* TMS
transitive inference problems, 234
trials, 79, 96, 114, 116–17, 172, 192, 234, 349, 374, 377, 418

uncertainty, 229, 244, 249–51, 253, 262, 268, 364, 366, 371, 387, 459
unconditioned stimulus, 79
underlying brain mechanisms, 17, 73
understanding of mind, 467, 471–72
upright letter, 88–89

Verbal Behavior, 206–7, 263
verb phrase. *See* VP
verbs, leg-action, 441
very large-scale integration. *See* VLSI
visual cortex, 90, 132, 135, 149, 151, 158, 393, 439, 456
visual field, 71, 87, 96, 98–99, 167, 169, 172, 213, 307, 315, 459–60
  right, 113, 160, 286
visual imagery, 18, 57, 111, 125–26, 132, 147, 150, 274, 439, 457
visual images, 41, 125, 128–29, 132, 145, 243, 286
visual system, 48, 86, 92–94, 96, 160–62, 178, 187, 190, 213–14, 365, 456–59
VLSI (very large-scale integration), 392–93
voluntarism, 57–58, 62, 64–66
VP (verb phrase), 271–72

weights, 46, 193–96, 199–200, 202, 226, 237, 372, 375, 382–83, 411, 414–15
  connection, 195–96, 202
  neuronal, 375
Wernicke, 289–92, 295
wholes, 58, 64, 69, 71, 73, 94, 143, 213
word meanings, 135, 200, 268–69, 440–41
words, 59–60, 90, 116, 160, 200–202, 204, 209–11, 213, 244, 246, 264–72, 274–75, 277–79, 281, 288–92, 306–7, 309–10, 324, 348–50, 440–41
  ambiguous, 270
  first, 204
  important, 103
  list of, 59–60, 101, 441
  short, 269
working memory, 114–17, 121–25, 145, 176, 178–80, 303, 393, 446, 469, 471
working memory function, 180, 471
working memory model, 118, 123–25, 145, 179–80
working memory system, 179–80
World Wide Web, 217