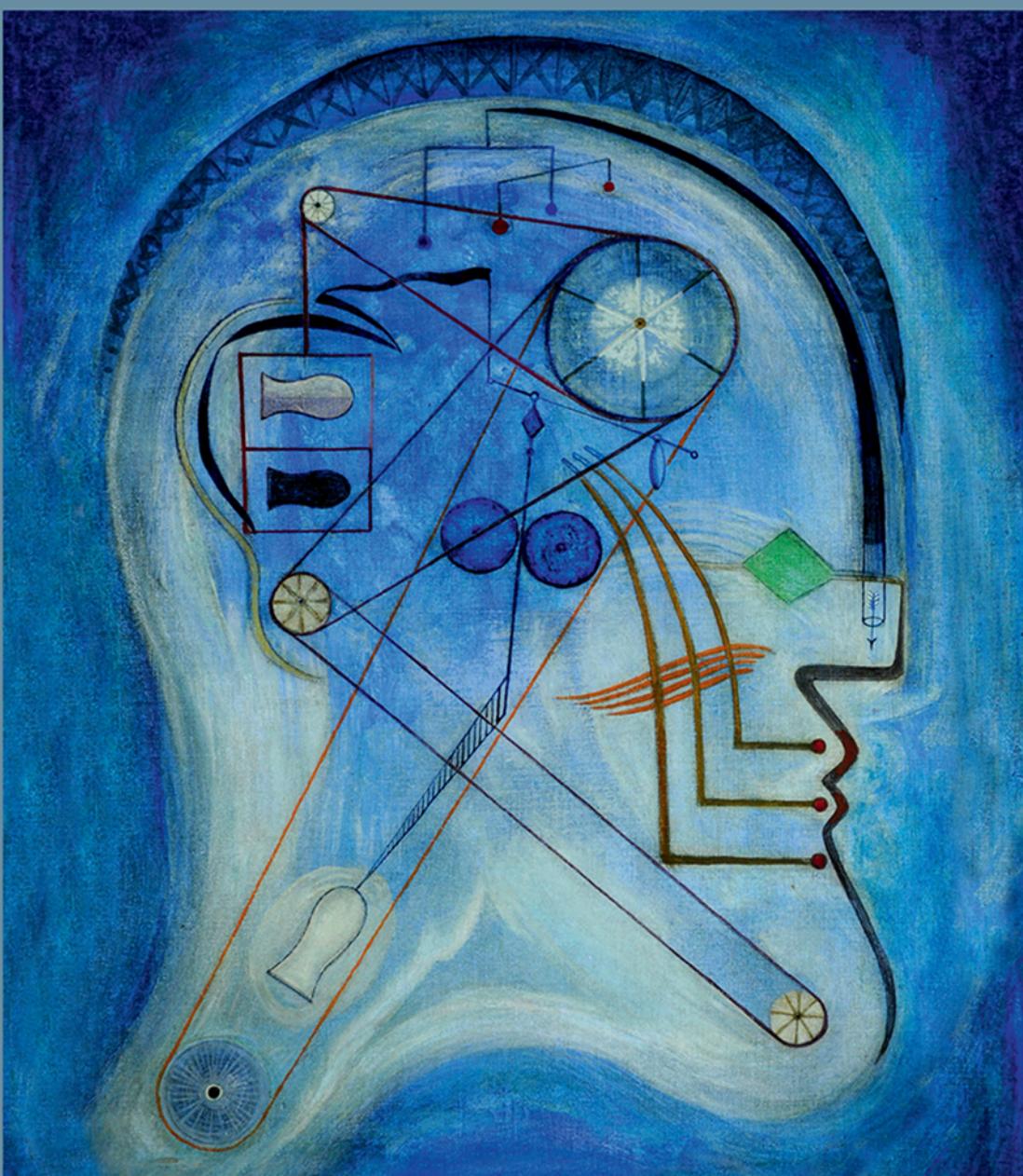


José Luis Bermúdez

Cognitive Science

An Introduction to the Science of the Mind

Third Edition

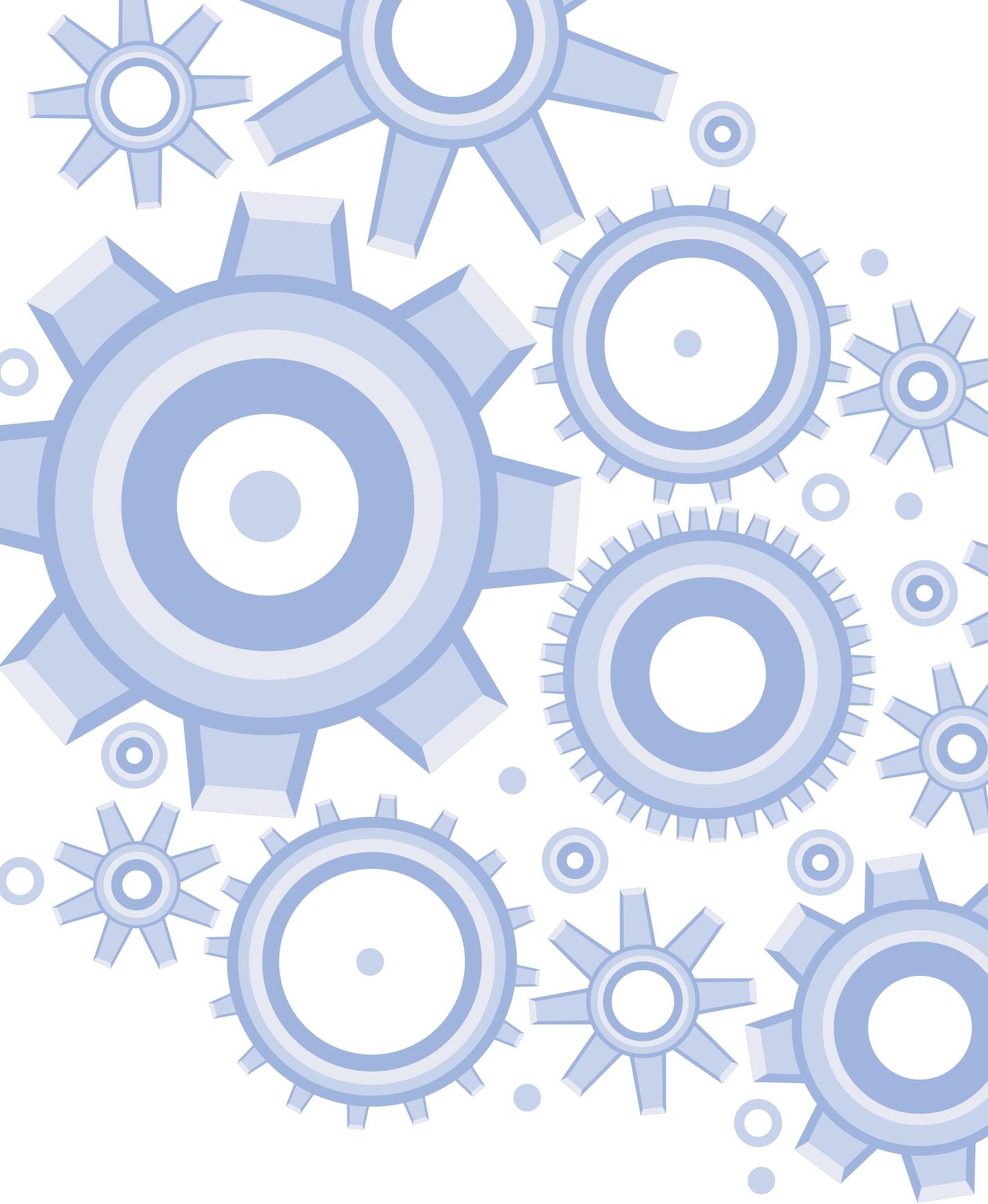




COGNITIVE SCIENCE

Third edition

The third edition of this popular and engaging text consolidates the interdisciplinary streams of cognitive science to present a unified narrative of cognitive science as a discipline in its own right. It teaches students to apply the techniques and theories of the cognitive scientist's "tool kit" – the vast range of methods and tools that cognitive scientists use to study the mind. Thematically organized, *Cognitive Science* underscores the problems and solutions of cognitive science rather than more narrowly examining individually the subjects that contribute to it – psychology, neuroscience, linguistics, and so on. The generous use of examples, illustrations, and applications demonstrates how theory is applied to unlock the mysteries of the human mind. Drawing upon cutting-edge research, the text has been substantially revised, with new material on Bayesian approaches to the mind and on deep learning. An extensive online set of resources is available to aid instructors and students alike. Sample syllabi show how the text can support a variety of courses, making it a highly flexible teaching and learning resource at both the undergraduate and graduate levels.



COGNITIVE SCIENCE

An Introduction to the Science of the Mind

Third edition

José Luis Bermúdez
Texas A&M University



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom
One Liberty Plaza, 20th Floor, New York, NY 10006, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India
79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781108424493

DOI: [10.1017/9781108339216](https://doi.org/10.1017/9781108339216)

© Cambridge University Press 2020

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2020

Printed in Singapore by Markono Print Media Pte Ltd

A catalogue record for this publication is available from the British Library.

ISBN 978-1-108-42449-3 Hardback

ISBN 978-1-108-44034-9 Paperback

Additional resources for this publication at www.cambridge.org/bermudez3e

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.



CONTENTS

List of Boxes *xiv*

List of Figures *xv*

List of Tables *xxii*

Preface *xxiii*

Acknowledgments for the First Edition *xxxii*

Acknowledgments for the Second Edition *xxxii*

Acknowledgments for the Third Edition *xxxiii*

Introduction: The Challenge of Cognitive Science *3*

PART I HISTORICAL LANDMARKS *12*

- 1** The Prehistory of Cognitive Science *15*
- 2** The Discipline Matures: Three Milestones *37*
- 3** The Turn to the Brain *65*

PART II MODELS AND TOOLS *96*

- 4** Physical Symbol Systems and the Language of Thought *99*
- 5** Neural Networks and Distributed Information Processing *123*
- 6** Applying Dynamical Systems Theory to Model the Mind *149*
- 7** Bayesianism in Cognitive Science *171*
- 8** Modules and Architectures *203*
- 9** Strategies for Brain Mapping *229*

PART III APPLICATIONS *256*

- 10** Models of Language Learning *259*
- 11** Object Perception and Folk Physics *285*
- 12** Machine Learning: From Expert Systems to Deep Learning *307*
- 13** Exploring Mindreading *335*
- 14** Mindreading: Advanced Topics *357*
- 15** The Cognitive Science of Consciousness *379*
- 16** Robotics: From GOFAI to Situated Cognition and Behavior-Based Robotics *407*
- 17** Looking Ahead: Challenges and Opportunities *437*

Glossary *444*

Bibliography *454*

Index for Cognitive Science (3rd edition) *478*



CONTENTS

List of Boxes *xiv*

List of Figures *xv*

List of Tables *xxii*

Preface *xxiii*

Acknowledgments for the First Edition *xxxii*

Acknowledgments for the Second Edition *xxxii*

Acknowledgments for the Third Edition *xxxiii*

Introduction: The Challenge of Cognitive Science *3*

0.1 Cognitive Science: An Interdisciplinary Endeavor *3*

0.2 Levels of Explanation: The Contrast between Psychology and Neuroscience *5*

How Psychology Is Organized *6*

How Neuroscience Is Organized *7*

0.3 The Challenge of Cognitive Science *10*

Three Dimensions of Variation *10*

The Space of Cognitive Science *10*

PART I HISTORICAL LANDMARKS *12*

1 The Prehistory of Cognitive Science *15*

1.1 The Reaction against Behaviorism in Psychology *16*

Learning without Reinforcement: Tolman and Honzik, “‘Insight’ in Rats” (1930) *17*

Cognitive Maps in Rats? Tolman, Ritchie, and Kalish, “Studies in Spatial Learning” (1946) *20*

Plans and Complex Behaviors: Lashley, “The Problem of Serial Order in Behavior” (1951) *21*

1.2 The Theory of Computation and the Idea of an Algorithm *22*

Algorithms and Turing Machines: Turing, “On Computable Numbers, with an Application to the Decision Problem” (1936–7) *23*

1.3 Linguistics and the Formal Analysis of Language *25*

The Structure of Language: Chomsky’s *Syntactic Structures* (1957) *26*

1.4 Information-Processing Models in Psychology	28
How Much Information Can We Handle? George Miller's "The Magical Number Seven, Plus or Minus Two" (1956)	29
The Flow of Information: Donald Broadbent's "The Role of Auditory Localization in Attention and Memory Span" (1954) and <i>Perception and Communication</i> (1958)	30
1.5 Connections and Points of Contact	32
2 The Discipline Matures: Three Milestones	37
2.1 Language and Micro-worlds	38
Natural Language Processing: Winograd, <i>Understanding Natural Language</i> (1972)	39
SHRDLU in Action	41
2.2 How Do Mental Images Represent?	47
Mental Rotation: Shepard and Metzler, "Mental Rotation of Three-Dimensional Objects" (1971)	48
Information Processing in Mental Imagery	50
2.3 An Interdisciplinary Model of Vision	53
Levels of Explanation: Marr's <i>Vision</i> (1982)	53
Applying Top-Down Analysis to the Visual System	55
3 The Turn to the Brain	65
3.1 Cognitive Systems as Functional Systems?	66
3.2 The Anatomy of the Brain and the Primary Visual Pathway	68
The Two Visual Systems Hypothesis: Ungerleider and Mishkin, "Two Cortical Visual Systems" (1982)	70
3.3 Extending Computational Modeling to the Brain	76
A New Set of Algorithms: Rumelhart, McClelland, and the PDP Research Group, <i>Parallel Distributed Processing: Explorations in the Microstructure of Cognition</i> (1986)	77
Pattern Recognition in Neural Networks: Gorman and Sejnowski, "Analysis of Hidden Units in a Layered Network Trained to Identify Sonar Targets" (1998)	78
3.4 Mapping the Stages of Lexical Processing	80
Functional Neuroimaging with PET: Petersen, Fox, Posner, and Mintun, "Positron Emission Tomographic Studies of the Cortical Anatomy of Single-Word Processing" (1988)	81
Petersen, Fox, Posner, and Mintun, "Positron Emission Tomographic Studies of the Cortical Anatomy of Single-Word Processing" (1988)	81
3.5 Studying Memory for Visual Events	84
Functional Neuroimaging with fMRI	86
Brewer, Zhao, Desmond, Glover, and Gabrieli, "Making Memories: Brain Activity That Predicts How Well Visual Experience Will Be Remembered" (1998)	87

3.6 The Neural Correlates of the BOLD Signal 90

Logothetis, "The Underpinnings of the BOLD Functional Magnetic Resonance Imaging Signal" (2001) 91

PART II MODELS AND TOOLS 96**4 Physical Symbol Systems and the Language of Thought 99****4.1 The Physical Symbol System Hypothesis 100**

Symbols and Symbol Systems 101

Transforming Symbol Structures 102

Intelligent Action and the Physical Symbol System 106

4.2 From Physical Symbol Systems to the Language of Thought 106

Intentional Realism and Causation by Content 108

The Language of Thought and the Relation between Syntax and Semantics 110

4.3 The Russian Room Argument and the Turing Test 114

Responding to the Russian Room Argument 117

5 Neural Networks and Distributed Information Processing 123**5.1 Neurally Inspired Models of Information Processing 124**

Neurons and Network Units 125

5.2 Single-Layer Networks and Boolean Functions 128

Learning in Single-Layer Networks: The Perceptron Convergence Rule 131

Linear Separability and the Limits of Perceptron Convergence 134

5.3 Multilayer Networks 137

The Backpropagation Algorithm 138

How Biologically Plausible Are Neural Networks? 139

5.4 Information Processing in Neural Networks: Key Features 141

Distributed Representations 141

No Clear Distinction between Information Storage and Information Processing 142

The Ability to Learn from "Experience" 143

6 Applying Dynamical Systems Theory to Model the Mind 149**6.1 Cognitive Science and Dynamical Systems 149**

What Are Dynamical Systems? 150

The Dynamical Systems Hypothesis: Cognitive Science without Representations? 153

6.2 Applying Dynamical Systems: Two Examples from Child Development 158

Two Ways of Thinking about Motor Control 159

Dynamical Systems and the A-Not-B Error 161

Assessing the Dynamical Systems Approach 166

7 Bayesianism in Cognitive Science	171
 7.1 Bayesianism: A Primer	172
Degrees of Belief and Subjective Probability	173
Conditional Probability	175
Bayes's Rule (the Short Version)	176
 7.2 Perception as a Bayesian Problem	179
The Predictive Challenge of Perception	179
Case Study: Binocular Rivalry	182
 7.3 Neuroeconomics: Bayes in the Brain	186
What Is Expected Utility?	187
Case Study: Neurons That Code for Expected Utility	190
<i>Probability-Detecting Neurons</i>	193
<i>Utility-Detecting Neurons</i>	194
<i>Combining Probability and Utility</i>	196
8 Modules and Architectures	203
 8.1 Architectures for Artificial Agents	204
Three Agent Architectures	204
 8.2 Fodor on the Modularity of Mind	208
Modular and Nonmodular Processing	208
 8.3 The Massive Modularity Hypothesis	210
The Cheater Detection Module	211
The Evolution of Cooperation	213
Two Arguments	216
Evaluating the Arguments for Massive Modularity	218
 8.4 Hybrid Architectures: The Example of ACT-R	219
The ACT-R Architecture	220
ACT-R as a Hybrid Architecture	222
9 Strategies for Brain Mapping	229
 9.1 Structure and Function in the Brain	230
Exploring Anatomical Connectivity	232
 9.2 Studying Cognitive Functioning: Techniques from Neuroscience	237
Mapping the Brain's Electrical Activity: EEG and MEG	237
Mapping the Brain's Blood Flow and Blood Oxygen Levels: PET and fMRI	240
 9.3 Combining Resources I: The Locus of Selection Problem	241
Combining ERPs and Single-Unit Recordings	242

9.4 Combining Resources II: Networks for Attention 246

Two Hypotheses about Visuospatial Attention 248

9.5 From Data to Maps: Problems and Pitfalls 249

From Blood Flow to Cognition? 250

Noise in the System? 251

Functional Connectivity versus Effective Connectivity 252

PART III APPLICATIONS 256**10 Models of Language Learning 259****10.1 Language and Rules 260**

Understanding a Language and Learning a Language 261

10.2 Language Learning and the Language of Thought: Fodor's Argument 263**10.3 Language Learning in Neural Networks 266**

The Challenge of Tense Learning 267

Connectionist Models of Tense Learning 269

10.4 Bayesian Language Learning 274

Probabilities in Word and Phrase Segmentation 275

Understanding Pronouns 276

Learning Linguistic Categories 278

11 Object Perception and Folk Physics 285**11.1 Object Permanence and Physical Reasoning in Infancy 286**

Infant Cognition and the Dishabituation Paradigm 286

How Should the Dishabituation Experiments Be Interpreted? 292

11.2 Neural Network Models of Children's Physical Reasoning 293

Modeling Object Permanence 295

Modeling the Balance Beam Problem 297

11.3 Conclusion: The Question of Levels 300**12 Machine Learning: From Expert Systems to Deep Learning 307****12.1 Expert Systems and Machine Learning 308**

Expert Systems and Decision Trees 308

ID3: An Algorithm for Machine Learning 310

12.2 Representation Learning and Deep Learning 315

Deep Learning and the Visual Cortex 318

12.3 The Machinery of Deep Learning 321

Autoencoders 322

	Convolutional Neural Networks	324
	<i>Sparse Connectivity</i>	325
	<i>Shared Weights</i>	326
	<i>Invariance under Translation</i>	326
	12.4 Deep Reinforcement Learning	327
13	Exploring Mindreading	335
	13.1 Pretend Play and Metarepresentation	336
	The Significance of Pretend Play	336
	Leslie on Pretend Play and Metarepresentation	337
	The Link to Mindreading	341
	13.2 Metarepresentation, Autism, and Theory of Mind	341
	Using the False Belief Task to Study Mindreading	342
	Interpreting the Results	344
	Implicit and Explicit Understanding of False Belief	347
	13.3 The Mindreading System	348
	First Steps in Mindreading	349
	From Dyadic to Triadic Interactions: Joint Visual Attention	351
	TESS and TOMM	352
14	Mindreading: Advanced Topics	357
	14.1 Why Does It Take Children So Long to Learn to Understand False Belief?	358
	Leslie's Answer: The Selection Processor Hypothesis	358
	An Alternative Model of Theory of Mind Development	360
	14.2 Mindreading as Simulation	363
	Standard Simulationism	363
	Radical Simulationism	365
	14.3 The Cognitive Neuroscience of Mindreading	365
	Neuroimaging Evidence for a Dedicated Theory of Mind System?	366
	Neuroscientific Evidence for Simulation in Low-Level Mindreading?	369
	Neuroscientific Evidence for Simulation in High-Level Mindreading?	373
15	The Cognitive Science of Consciousness	379
	15.1 The Challenge of Consciousness: The Knowledge Argument	380
	15.2 Information Processing without Conscious Awareness: Some Basic Data	382
	Consciousness and Priming	382
	Nonconscious Processing in Blindsight and Unilateral Spatial Neglect	384

15.3 So What Is Consciousness For?	387
What Is Missing in Blindsight and Spatial Neglect	389
Milner and Goodale: Vision for Action and Vision for Perception	389
What Is Missing in Masked Priming	392
15.4 Two Types of Consciousness and the Hard Problem	393
15.5 The Global Workspace Theory of Consciousness	396
The Building Blocks of Global Workspace Theory	396
The Global Neuronal Workspace Theory	397
15.6 Conclusion	400
16 Robotics: From GOFAI to Situated Cognition and Behavior-Based Robotics	407
16.1 GOFAI Robotics: SHAKEY	408
SHAKEY's Software I: Low-Level Activities and Intermediate-Level Actions	409
SHAKEY's Software II: Logic Programming in STRIPS and PLANEX	413
16.2 Situated Cognition and Biorobotics	414
The Challenge of Building a Situated Agent	415
Situated Cognition and Knowledge Representation	416
Biorobotics: Insects and Morphological Computation	418
16.3 From Subsumption Architectures to Behavior-Based Robotics	423
Subsumption Architectures: The Example of Allen	424
Behavior-Based Robotics: TOTO	427
Multiagent Programming: The Nerd Herd	430
17 Looking Ahead: Challenges and Opportunities	437
17.1 Exploring the Connectivity of the Brain: The Human Connectome Project and Beyond	438
17.2 Understanding What the Brain Is Doing When It Appears Not to Be Doing Anything	439
17.3 Neural Prosthetics	440
17.4 Cognitive Science and the Law	441
17.5 Autonomous Vehicles: Combining Deep Learning and Intuitive Knowledge	442
<i>Glossary</i>	444
<i>Bibliography</i>	454
<i>Index for Cognitive Science (3rd edition)</i>	478



BOXES

- 2.1** A Conversation with ELIZA *39*
- 3.1** What Does Each Lobe Do? *69*
- 3.2** Brain Vocabulary *72*
- 4.1** Defining Sentences in Propositional Logic *102*
- 6.1** Basins of Attraction in State Space *157*
- 7.1** Basic of the Probability Calculus *174*
- 7.2** Deriving Bayes's Rule *177*
- 8.1** The Prisoner's Dilemma *215*
- 15.1** A Typical Semantic Priming Experiment *384*



FIGURES

- 0.1** Connections among the cognitive sciences, as depicted in the Sloan Foundation's 1978 report. [4](#)
- 0.2** Some of the principal branches of scientific psychology. [7](#)
- 0.3** Levels of organization and levels of explanation in the nervous system. [8](#)
- 0.4** The spatial and temporal resolution of different tools and techniques in neuroscience. [9](#)
- 0.5** The “space” of contemporary cognitive science. [11](#)
- 1.1** A rat in a Skinner box. [18](#)
- 1.2** A fourteen-unit T-Alley maze. [19](#)
- 1.3** A cross-maze. [20](#)
- 1.4** Schematic representation of a Turing machine. [25](#)
- 1.5** A sample phrase structure tree for the sentence “John has hit the ball.” [27](#)
- 1.6** Donald Broadbent’s (1958) model of selective attention. [29](#)
- 2.1** A question for SHRDLU about its virtual micro-world. [40](#)
- 2.2** An algorithm for determining whether a given input is a sentence or not. [42](#)
- 2.3** Algorithms for identifying noun phrases and verb phrases. [43](#)
- 2.4** Procedure for applying the command CLEARTOP. [44](#)
- 2.5** SHRDLU acting on the initial command to pick up a big red block. [45](#)
- 2.6** SHRDLU completing instruction 3 in the dialog: “Find a block which is taller than the one you are holding and put it in the box.” [46](#)
- 2.7** Examples of the three-dimensional figures used in Shepard and Metzler’s 1971 studies of mental rotation. [48](#)
- 2.8** Results of Shepard and Metzler’s 1971 studies of mental rotation. [49](#)
- 2.9** Examples of vertically and horizontally oriented objects that subjects were asked to visualize in Kosslyn’s 1973 scanning study. [52](#)
- 2.10** Two images of a bucket. [56](#)

- 2.11** Two examples of Marr's primal sketch, the first computational stage in his analysis of the early visual system. [57](#)
- 2.12** An example of part of the 2.5D sketch. [58](#)
- 2.13** An illustration of Marr's 3D sketch, showing how the individual components are constructed. [59](#)
- 2.14** The place of the implementational level within Marr's overall theory. [60](#)
- 2.15** An illustration of the hierarchical organization of the visual system, including which parts of the brain are likely responsible for processing different types of visual information. [61](#)
- 3.1** The large-scale anatomy of the brain, showing the forebrain, the midbrain, and the hindbrain. [69](#)
- 3.2** A vertical slice of the human brain, showing the cerebrum. [70](#)
- 3.3** The division of the left cerebral hemisphere into lobes. [71](#)
- 3.4** The primary visual pathway. [72](#)
- 3.5** Image showing ventral (purple) and dorsal (green) pathways in the human visual system. [73](#)
- 3.6** Design and results of Ungerleider and Mishkin's cross-lesion disconnection studies. [75](#)
- 3.7** A generic three-layer connectionist network (also known as an artificial neural network). [78](#)
- 3.8** Gorman and Sejnowski's mine/rock detector network. [80](#)
- 3.9** Images showing the different areas of activation (as measured by blood flow) during the four different stages in Petersen et al.'s (1988) lexical access studies. [84](#)
- 3.10** A flowchart relating areas of activation to different levels of lexical processing. [85](#)
- 3.11** Neural area showing activity when subjects looked at pictures. [88](#)
- 3.12** Neural areas where activation is correlated with levels of memory performance. [89](#)
- 3.13** A microelectrode making an extracellular recording. [90](#)
- 3.14** Simultaneous microelectrode and fMRI recordings from a cortical site showing the neural response to a pulse stimulus of 24 seconds. [92](#)
- 4.1** A typical traveling salesperson problem. [104](#)
- 4.2** The structure of Fodor's argument for the language of thought hypothesis. [114](#)
- 4.3** Inside and outside the Russian room. [116](#)
- 5.1** Schematic illustration of a typical neuron. [125](#)

- 5.2** An artificial neuron. 126
- 5.3** Four different activation functions. 127
- 5.4** Illustration of a mapping function. 128
- 5.5** A single-layer network representing the Boolean function AND. 130
- 5.6** A single-layer network representing the Boolean function NOT. 131
- 5.7** The starting configuration for a single-layer network being trained to function as a NOT-gate through the perceptron convergence rule. 133
- 5.8** Graphical representations of the AND and XOR (exclusive-OR) functions, showing the linear separability of AND. 135
- 5.9** A multilayer network representing the XOR (exclusive-OR) function. 136
- 5.10** The computational operation performed by a unit in a connectionist model. 138
- 6.1** The trajectory through state space of an idealized swinging pendulum. 151
- 6.2** The state space of a swinging pendulum in a three-dimensional phase space. 152
- 6.3** Illustration of the Watt governor, together with a schematic representation of how it works. 155
- 6.4** An example of the computational approach to motor control. 160
- 6.5** The stage IV search task, which typically gives rise to the A-not-B-error in infants at around the age of 9 months. 162
- 6.6** An infant sitting for an A trial and standing for a B trial. 163
- 6.7** Applying the dynamical field model to the A-not-B error. 165
- 7.1** An illustration purporting to be of Thomas Bayes from a 1936 book on the history of life insurance. 172
- 7.2** A diagram showing the proportion of the probability space in which A is true; the proportion of the probability space in which B is true; and the intersection of A and B (which is the region where A and B are both true). 175
- 7.3** Four of the seven Gestalt principles of grouping, illustrated and explained. 180
- 7.4** Two examples of stimuli used to elicit binocular rivalry. 182
- 7.5** Two well-known ambiguous figures: Rubin's vase and the duck–rabbit illusion. 183
- 7.6** The principal pathways for saccade production. 192
- 7.7** Platt and Glimcher's probabilistic cued saccade task. 193
- 7.8** Activity of an LIP neuron during the probability experiment. 194

- 7.9** Platt and Glimcher's cued saccade experiment, with stimulus and response held constant and the quantity of reward varied. 195
- 7.10** Activity of an LIP neuron while a monkey makes his own choice compared to a behaviorally derived estimate of the value of the movement to the monkey. 197
- 8.1** The architecture of a simple reflex agent. 205
- 8.2** The architecture of a goal-based agent. 206
- 8.3** The architecture of a learning agent. 207
- 8.4** A version of the Wason selection task. 212
- 8.5** A version of Griggs and Cox's deontic selection task. 213
- 8.6** The evolutionary biologist W. D. Hamilton (1936–2000). 217
- 8.7** The ACT-R cognitive architecture. 221
- 9.1** Luria's (1970) diagram of the functional organization of the brain. 231
- 9.2** Map of the anatomy of the brain showing the four lobes and the Brodmann areas. 233
- 9.3** A connectivity matrix for the visual system of the macaque monkey. 235
- 9.4** An anatomical wiring diagram of the visual system of the macaque monkey. 236
- 9.5** The results of single-neuron recordings of a mirror neuron in area F5 of the macaque inferior frontal cortex. 238
- 9.6** Typical patterns of EEG waves, together with where/when they are typically found. 239
- 9.7a** Common experimental design for neurophysiological studies of attention. 243
- 9.7b** Example of the occipital ERPs recorded in a paradigm of this nature. 244
- 9.7c** Single-unit responses from area V4 in a similar paradigm. 245
- 9.7d** Single-unit responses from area V1 showing no effect of attention. 245
- 9.8** Frontoparietal cortical network during peripheral visual attention. 247
- 9.9** An illustration of a typical delayed saccade task. 248
- 9.10** Peripheral attention versus spatial working memory versus saccadic eye movement across studies. 250
- 10.1** The dual-route model of past tense learning in English proposed by Steven Pinker and Alan Prince. 269
- 10.2** Rumelhart and McClelland's model of past tense acquisition. 270
- 10.3** Performance data for Rumelhart and McClelland's model of past tense learning. 271

-
- 10.4 The network developed by Plunkett and Marchman to model children's learning of the past tense. 272
 - 10.5 A comparison of the errors made by Adam, a child studied by the psychologist Gary Marcus, and the Plunkett–Marchman neural network model of tense learning. 273
 - 10.6 A hierarchical cluster of similarity judgments, with nodes corresponding to clusters of stimuli more similar on average to each other than to objects in the nearest cluster. 279
 - 11.1 Schematic representation of the habituation and test conditions in Baillargeon's drawbridge experiments. 288
 - 11.2 Schematic representation of an experiment used to test infants' understanding of object boundaries and sensitivity to what Spelke calls the principle of cohesion. 289
 - 11.3 Schematic representation of an experiment testing infants' understanding of the principle of contact. 290
 - 11.4 Schematic depiction of events that accord with, or violate, the continuity or solidity constraints. 291
 - 11.5 A series of inputs to the network as a barrier moves in front of a ball and then back to its original location. 295
 - 11.6 Recurrent network for learning to anticipate the future position of objects. 296
 - 11.7 A balance beam. 297
 - 11.8 The architecture of the McClelland and Jenkins network for the balance beam problem. 299
 - 12.1 A decision tree illustrating a mortgage expert system. 309
 - 12.2 The first node on the decision tree for the tennis problem. 313
 - 12.3 The complete decision tree generated by the ID3 algorithm. 313
 - 12.4 A sample completed questionnaire used as input to an ID3-based expert system for diagnosing diseases in soybean crops. 314
 - 12.5 Different ways of distinguishing two groups in a database of examples. 317
 - 12.6 An illustration of hierarchical visual processing. 320
 - 12.7 Illustration of how an autoencoder compresses and then decompresses a signal. 323
 - 12.8 A move in the Google DeepMind challenge between AlphaGo and Lee Sedol in 2016. 328
 - 13.1 An example of metarepresentation. 338
 - 13.2 The general outlines of Leslie's model of pretend play. 339
 - 13.3 Leslie's Decoupler model of pretense. 340

- 13.4** The task used by Baron-Cohen, Leslie, and Frith to test for children's understanding of false belief. [344](#)
- 13.5** Illustration of the connection between pretend play and success on the false belief task. [346](#)
- 13.6** Baron-Cohen's model of the mindreading system. [350](#)
- 14.1** What goes on when one subject represents another's belief. [361](#)
- 14.2** What goes on when one subject represents another's perception. [362](#)
- 14.3** A schematic version of standard simulationism. [364](#)
- 14.4** Schematic representation of brain regions associated with the attribution of mental states. [367](#)
- 14.5** Schematic overview of the frontoparietal mirror neuron system (MNS) and its main visual input in the human brain. [372](#)
- 15.1** An illustration of a typical priming experiment. [382](#)
- 15.2** Examples of deficits found in patients with left spatial neglect (damage to the right hemisphere of the brain). [386](#)
- 15.3** D.B.'s responses to pictures of animals presented in his blind field. [388](#)
- 15.4** An illustration of the two houses presented to P.S. [389](#)
- 15.5** In this experiment, subjects were asked either to "post" a card into a slot or to rotate another hand-held card to match the orientation of the slot. [391](#)
- 15.6** In the Ebbinghaus illusion, two circles are illusorily seen as differently sized, depending on what surrounds them. [392](#)
- 15.7** In the Norman and Shallice 1980 model, conscious processing is involved in the supervisory attentional regulation, by prefrontal cortices, of lower-level sensorimotor chains. [398](#)
- 15.8** The neural substrates of the global workspace. [399](#)
- 16.1** A map of SHAKEY's physical environment. [409](#)
- 16.2** A photograph of SHAKEY the robot. [410](#)
- 16.3** The organizing principles of biorobotics – a highly interdisciplinary enterprise. [419](#)
- 16.4** The cricket's ears are on its front legs. [420](#)
- 16.5** A robot fish called WANDA. [421](#)
- 16.6** WANDA swimming upward. [422](#)

- 16.7** Another example of morphological computation: the robot hand designed by Hiroshi Yokoi. [422](#)
- 16.8** The Yokoi hand grasping two very different objects. [423](#)
- 16.9** Rodney Brooks's robot Allen, his first subsumption architecture robot. [424](#)
- 16.10** The layers of Allen's subsumption architecture. [425](#)
- 16.11** The Nerd Herd, together with the pucks that they can pick up with their grippers. [430](#)



TABLES

- 2.1** A table illustrating the three different levels that Marr identified for explaining information-processing systems [55](#)
- 4.1** Syntax and semantics in the predicate calculus [113](#)
- 8.1** Comparing the symbolic and subsymbolic dimensions of knowledge representation in the hybrid ACT-R architecture [224](#)
- 9.1** Comparing techniques for studying connectivity in the brain [241](#)
- 10.1** The stages of past tense learning according to verb type [268](#)
- 13.1** The three groups studied in Baron-Cohen, Leslie, and Frith (1985) [343](#)
- 16.1** SHAKEY'S five levels [411](#)
- 16.2** How SHAKEY represents its own state [412](#)
- 16.3** SHAKEY's intermediate-level routines [413](#)
- 16.4** The five basis behaviors programmed into Matarić's Nerd Herd robots [431](#)



PREFACE



About This Book

There are few things more fascinating than the human mind – and few things that are more difficult to understand. Cognitive science is the enterprise of trying to make sense of this most complex and baffling natural phenomenon.

The very things that make cognitive science so fascinating make it very difficult to study and to teach. Many different disciplines study the mind. Neuroscientists study the mind's biological machinery. Psychologists directly study mental processes, such as perception and decision-making. Computer scientists explore how those processes can be simulated and modeled in computers. Evolutionary biologists and anthropologists speculate about how the mind evolved. In fact, very few academic areas are not relevant to the study of the mind in some way. The job of cognitive science is to provide a framework for bringing all these different perspectives together.

The enormous range of information out there about the mind can be overwhelming, both for students and for instructors. Different textbooks have approached this challenge in different ways.

Some textbooks have concentrated on being as comprehensive as possible, with a chapter covering key ideas in each of the relevant disciplines – a chapter on psychology, a chapter on neuroscience, and so on. These books are often written by committee – with each chapter written by an expert in the relevant field. These books can be very valuable, but they really give an introduction to the cognitive sciences (in the plural) rather than to cognitive science as an interdisciplinary enterprise.

Other textbook writers take a much more selective approach, introducing cognitive science from the perspective of the disciplines that they know best – from the perspective of philosophy, for example, or of computer science. Again, I have learned much from these books, and they can be very helpful. But I am convinced that students and instructors need something more general.

This book aims for a balance between these two extremes. Cognitive science has its own problems and its own theories. The book is organized around these. They are all ways of working out the fundamental idea at the heart of cognitive science – which is that the mind is an information processor. What makes cognitive science so rich is that this single basic idea can be (and has been) worked out in many different ways. In presenting these different models of the mind as an information processor, I have tried to select as

wide a range of examples as possible to give students a sense of cognitive science's breadth and range.



About the Third Edition

Cognitive Science: An Introduction to the Science of the Mind has been very significantly revised for the third edition. These changes have been made for two reasons. First, I wanted to make the book more accessible to students in the first and second years of their studies. To achieve that goal, I have made changes to both content and style as well as to the organization of the book. Second, I wanted to make sure that the new edition reflects the most exciting new developments in cognitive science, some of which were barely discernible back in 2010, when the first edition was published.

Previous editions of this book were organized around what I termed the *integration challenge*. This is the challenge of providing a unified theoretical framework for studying cognition that brings together the different disciplines that study the mind. The third edition no longer uses the integration challenge as an organizing principle. The additional layer of complexity is useful for many purposes, but not, I now think, for pedagogical ones. As a result, I have cut the two chapters that were devoted to the integration challenge in the first and second editions and simplified the presentation in later chapters. In particular, I no longer employ a two-way division into symbolic and nonsymbolic models of information processing. I have added an introduction that explains in some more general terms some of the issues and problems previously discussed under the label “integration challenge.”

I have used the space freed up by reorganization to expand coverage of more up-to-date areas elsewhere in the book. This includes a new chapter on Bayesian approaches to the mind. This chapter covers both the idea that cognition can be understood in terms of Bayesian hypothesis testing and error minimization and experimental studies in neuroeconomics of how probabilities and values seem to be calculated in a broadly Bayesian manner in the primate nervous system. In addition, I have updated the discussion of machine learning in what is now [Chapter 12](#), eliminating some by now dated examples and replacing them with more topical discussion of *deep learning* algorithms.

To help instructors and students, I have divided some of the longer chapters from the second edition into two. Dynamical systems theory has its own chapter ([Chapter 6](#)), while situated cognition and robotics are now in [Chapter 16](#). The lengthy discussion of mindreading in the second edition is now spread over two chapters: “Exploring Mindreading” ([Chapter 13](#)) and “Mindreading: Advanced Topics” ([Chapter 14](#)).



How the Book Is Organized

This book is organized into three parts.



Part I: Historical Landmarks

Cognitive science has evolved considerably in its short life. Priorities have changed as new methods have emerged – and some fundamental theoretical assumptions have changed with them. The three chapters in [Part I](#) introduce students to some of the highlights in the history of cognitive science. Each chapter is organized around key discoveries and/or theoretical advances.



Part II: Models and Tools

[Part II](#) sets out the main models and tools that cognitive scientists can bring to bear to understand cognition and the mind.

The first model, discussed in [Chapter 4](#), is associated with the physical symbol system hypothesis originally developed by the computer scientists Allen Newell and Herbert Simon. According to the physical symbol system hypothesis, all information processing involves the manipulation of physical structures that function as symbols. For the first decades of cognitive science, the physical symbol systems hypothesis was, as Jerry Fodor famously put it, the “only game in town.” In the 1980s and 1990s, connectionist and neural network modelers developed an alternative, derived from models of artificial neurons in computational neuroscience and connectionist artificial intelligence. [Chapter 5](#) explores the motivation for this approach and introduces some of the key concepts.

Another set of models and tools that can be used to study the mind derives from dynamical systems theory and is introduced and discussed in [Chapter 6](#). Bayesian approaches to modeling the mind have also gained currency. As explained in [Chapter 7](#), these approaches treat the mind as a predictive, hypothesis-testing machine and have been used both to study the mind as a whole and to model the activity of individual brain areas and populations of neurons.

One of the key ideas of cognitive science is that the mind is modular (that some, or all, information processing is carried out by specialized modules). [Chapter 8](#) explores different ways of developing this basic idea, including the radical claim, proposed by evolutionary psychologists, that the mind is simply a collection of specialized modules, with no non-specialized processing at all. Theoretical discussions of modularity are complemented by experimental techniques for studying the organization of the mind. [Chapter 9](#) surveys the cognitive scientist’s tool kit in this regard, focusing in particular on different types of brain mapping.



Part III: Applications

The seven chapters in this part are more applied than those in [Part II](#). They explore different ways in which the models and tools introduced in [Part II](#) can be used to give accounts of particular cognitive phenomena.

[Chapter 10](#) considers language learning. Many cognitive scientists have thought that language learning is a process of learning explicit rules with a significant innate component. But we explore both neural network and Bayesian models that illustrate alternative ways of thinking about how children can learn languages. In [Chapter 11](#) we turn to models of how children learn about the basic structure of the physical world (how they acquire what is often called a *folk physics*). Here, too, we see the power of neural network models.

One of the most significant recent developments of neural network models has been the explosive growth of deep learning algorithms, which have made possible impressive advances in areas long thought to be major challenges for artificial intelligence, such as image recognition, machine translation, and games of strategy, such as Go. These are covered in [Chapter 12](#).

[Chapters 13](#) and [14](#) illustrate how theoretical, methodological, and experimental issues can come together. They work through an issue that has received much attention in contemporary cognitive science – the issue of whether there is a dedicated cognitive system response for our understanding of other people (the so-called mindreading system). [Chapter 13](#) presents some of the basic issues and developments, while more advanced topics are introduced in [Chapter 14](#).

In [Chapter 15](#) we look at recent developments in the cognitive science of consciousness – a fast-moving and exciting area that raises fundamental questions about possible limits to what can be understood through the tools and techniques of cognitive science. And then finally, in [Chapter 16](#), we explore the situated cognition movement and related developments in robotics, particularly behavior-based robotics and biologically inspired robotics.



Using This Book in Courses

This book has been designed to serve as a self-contained text for a single-semester (12–15 weeks) introductory course on cognitive science. Students taking this course may have taken introductory courses in psychology and/or philosophy, but no particular prerequisites are assumed. All the necessary background is provided for a course at the freshman or sophomore level (first or second year). The book could also be used for a more advanced introductory course at the junior or senior level (third or fourth year). In this case, the instructor would most likely want to supplement the book with additional readings. There are suggestions on the instructor website (see below).



Text Features

I have tried to make this book as user-friendly as possible. Key text features include the following:

- **Chapter overviews.** Each chapter begins with an overview to orient the reader.

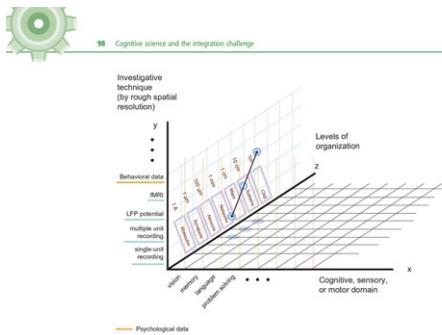


Figure 4.5 The integration challenge and the “space” of contemporary cognitive science.

In any event, whether the integration challenge is ultimately soluble or not, it is very clear that, as things stand, we are nowhere near solving it. Even the most ambitious theories and studies that have been carried out by cognitive scientists set out to cover only a small slice of the space of cognitive science. For example, Marr’s theory of vision is one of the more ambitious undertakings of cognitive science, and Marr’s tri-level hypothesis is often taken as a textbook example of how cognitive science can span different levels of explanation. But the target of Marr’s theory is really just a very small part of vision. Marr’s theory of vision is ultimately a theory of visual processing. It has nothing to say about object recognition and object identification, nor about how vision is organized when sensory input comes from both eyes. Information is stored in memory, so Marr’s theory of vision is really a very small slice of behavior, and thus of the *space* of cognitive science – alternatively, it occupies only a very small horizontal slice of cognitive science. Moving to the *x*-axis, Marr had relatively little to say about what he called the implementational level. And in fact, as we shall see in the next chapter (section 5.2), the very idea that there is a single implementational level is deeply flawed.

4.4 Evolutionary psychology

Local integration I: Evolutionary psychology and the psychology of reasoning

Cognitive psychologists have paid close attention to human problem-solving. We have already seen an example of this in the experiments on mental imagery and mental rotation. Another area of particular interest is the psychology of reasoning. The logic rotation approach – performed involving the inference of two figures, for example, when more attention has been paid to problems that are linguistically framed, such as problems where subjects have to determine how likely it is that certain propositions are true, or whether one proposition follows from (is entailed by) another. These problems are all reasoning problems, and psychologists have studied them with the aim of uncovering the mechanics of reasoning.

A second hypothesis that cognitive psychologists have come up with is that, through cues on logic and probability theory, is that human reasoning is governed by the basic principles of logic and probability theory. People exploit the basic principles of logic when they are trying to solve problems that have a determinate “yes/no” answer fixed by logical relations between propositions, and they use the principles of probability theory when the problem is to work out how likely some event is to happen. This may seem too obvious to be worth stating. How could we use anything but logic to solve logic problems? And how could we use anything but probability theory to solve probability problems?

– Actually, however, the hypothesis is far from obviously true. Logic and probability theory are branches of mathematics, not of psychology. They study abstract mathematical relations. Those abstract mathematical relations determine the correct solution to particular problems. But logic and probability theory have nothing to say about how we actually go about solving those problems. In order to work out the reasoning principles that we use, psychologists have to conduct experiments to work out the sorts of problems that we use, and to work out the sorts of problems that we do not use.

Before going on to look at some of those experiments we need to make explicit an important feature of both logic and probability theory. The basic laws of logic and principles of probability theory are universal. Logical relations hold between sentences irrespective of what those sentences actually say. We might, for example, make the following inference: “If that’s the cathedral, then the library must be over there. But it’s not. So, it’s not the cathedral.” The logical rule here is known as *modus tollens*. The rule is the same that a conditional (*If A then B*) and the negation of the consequent of that conditional (*not-B*) jointly entail the negation of the antecedent of that conditional (*not-A*).

In our example the sentence “that’s the cathedral” takes the place of *A* the antecedent of the conditional and “the library must be over there” takes the place of *B* the consequent of the conditional. What is distinctive about this sort of inference is that it makes no difference what sentences one puts in place of *A* and *B* in the standard terminology; this inferential transition is *domain-general*. Whatever one puts in place of *A* and *B* the

- **Exercises.** These have been inserted at various points within each chapter. They are placed in the flow of the text to encourage the reader to take a break from reading and engage with the material. They are typically straightforward, but for a few, I have placed suggested solutions on the instructor website (see below).



of representational primitives and possible parameters of variation. Once again, it is easy to see why informational encapsulation will secure computational tractability. An informationally encapsulated module with have only a limited range of inputs on which to work.

In contrast, modular processing runs very quickly into versions of the so-called *frame problem*. This is the problem, particularly pressing for those developing expert systems with AI and designing robots, of building into a system rules that will correctly identify what information and which inferences should be pursued in a given situation. The problem is identifying what sort of information is relevant and hence needs to be taken into account. Daniel Dennett’s classic article on the subject opens with the following amusing and instructive tale:

Once upon a time there was a robot, named R1 by its creators. Its only task was to find for itself. One day its designers arranged for it to learn that its spare battery, its precious energy supply, was locked in a room with a time bomb set to go off soon. R1 located the room, and the key to the door, and formulated a plan to rescue its battery. There was a wagon in the room, and R1 knew that the wagon was on the floor, and R1 knew that it was a certain action which it called PULLOUT (Wagon, Room), that would result in the battery being removed from the room. Straightaway it acted, and did succeed in getting the battery out of the room before the bomb went off. Unfortunately, however, the bomb was also on the wagon. R1 knew that the bomb was on the wagon in the room, but didn’t realize that pulling the wagon would bring the bomb out along with the battery. Poor R1 had to learn that lesson the hard way.

Back to the drawing board. “The solution is obvious,” said the designers. “Our next robot must be made to recognize not just the intended implications of its acts, but also the implications about their side-effects, by deducing these implications from the descriptions it uses in formulating its plans.” They called their next model, the robot-decoder, R1D1. They placed R1D1 in much the same predicament that R1 had succumbed to, and as it too hit upon the idea of PULLOUT (Wagon, Room), it began, as designed, to consider the implications of its actions. It deduced that pulling the wagon would bring the bomb out of the room would not change the colour of the room’s walls, and it embarked on a proof of the further implication that pulling the wagon out would cause its wheels to turn more revolutions than there were wheels on the wagon – when the bomb exploded.

Back to the drawing board. “We must teach it the difference between relevant implications and irrelevant implications,” said the designers, “and teach it to ignore the irrelevant ones.” So they developed a range of tagging schemes, such as either relevant or irrelevant, or even as relevant or irrelevant, and then they programmed the robot-decoder, or R2D1 for short. When they subjected R2D1 to the test that had so unequivocally selected its ancestors for extinction, they were surprised to see it sitting, Hamlet-like, outside the room containing the ticking bomb, the native hue of its resolution sickled o’er with the pale cast of thought, as Shakespeare (and more recently Fodor) has aptly put it. “Do something!” they yelled at it. “I am,” it retorted. “I’m busily ignoring some thousands of implications I have determined to be irrelevant. Just as soon

5.3 Models of mental architecture



as I find an irrelevant implication, I put it on the list of those I must ignore, and... * the bomb went off.

The greater the range of potentially relevant information, the more intractable this problem will be. This means that the tractability of the frame problem is in *inverse* proportion to the degree of information encapsulation. The more informationally encapsulated an informational system is, the less significant the frame problem will be. In the case of strictly modular systems, the frame problem will be negligible. In contrast, the less informationally encapsulated a system is, the more significant the frame problem will be. For non-modular systems, the frame problem has proven very hard indeed to tackle.

Exercise 5.7 Explain in your own words what the frame problem is, without reference to the robot example. Distinguish the three approaches to the problem that Dennett identifies in this passage (again without reference to the robot example) and explain the difficulty with each of them.

For these two reasons, then, it looks very much as if the type of top-down, algorithmic approach advocated by Marr works best for cognitive systems that are specialized, domain-specific and narrowly focused. This is not to say that for other cognitive systems it could not be extended to systems that are non-modular. Marr’s approach would still not be applicable to the mind as a whole. Whether or not it is possible to provide a functional specification susceptible to algorithmic formulation for high-level cognitive systems, it is hard to imagine what a functional specification would look like for the mind as a whole. But in the last analysis an understanding of the mind as a whole is what a solution to the integration challenge is ultimately aiming at.

5.3 Models of mental architecture

In this section we explore an alternative approach to the integration challenge – one that provides a much better fit with what is actually going on in contemporary cognitive science than either of the two global approaches we have been considering. The inter-theoretic reduction approach and the tri-level hypothesis both tackle the integration problem head-on. They take very seriously the idea that cognitive science spans different levels of organization. The approach we will be exploring in this section tackles the problem from a different direction. It starts off from a basic assumption common to all the cognitive sciences and then shows how different ways of interpreting that basic assumption generate different models of the mind as a whole. These different models of the mind as a whole are what I am calling different *mental architectures*. Each mental architecture is a way of unifying the different components and levels of cognitive science.

- **Boxes.** Boxes have been included to provide further information about the theories and research discussed in the text. Readers are encouraged to work through these, but the material is not essential to the flow of the text.



174 Applying the symbolic paradigm

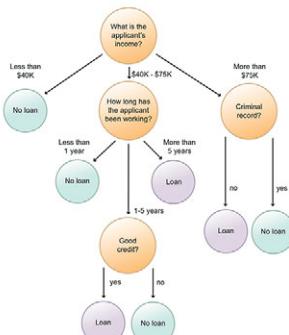


Figure 7.1 A decision tree illustrating a mortgage expert system. (From Friedenberg and Silverman 2006)

leads to a unique outcome (which computer scientists call a *terminal leaf* or node). But how are we supposed to get to these questions? How does the decision tree get designed, as it were?

One simple way of doing it would be to ask a team of mortgage loan officers to sit down and work out a decision tree that would more or less map onto the practices at their bank. This approach would be too slow and too expensive for most available programming languages. This would be fine, and it is no doubt that many expert systems programs are actually written (particularly in the mortgage area). But from the perspective of AI this would not be very interesting. It would be an expert system only in a very derivative sense. The real expert system would be the team of mortgage

7.1 Expert systems and machine learning

175



loan professionals. Much more interesting would be a program that was capable of producing its own decision tree – a program capable of imposing its own structure upon the problem and working out what would count as a solution. How would this be done?

Here is a more precise way of characterizing the problem. Suppose that we have a huge database of all the loan decisions that the bank has taken over a long period of time, together with all the relevant information about the applicants – their income, work history, credit rating, and so on. If we can find a way of representing the bank's past decisions in the form of a decision tree, so that each branch of the tree ends either in the loan being given or the loan being declined, then we can use that decision tree to ‘process’ new applications.

Machine learning and the physical symbol system hypothesis

We can put the same point the other way around. The decision tree in Figure 7.1 is a tool for analyzing new loan applications. The information that any applicant provides in response to the questions that the tree poses will steer the applicant down one of the branches and the applicant will end up with their application either being approved or turned down. So the challenge for the expert system is to come up with a decision tree that can in effect take a new set of pre-set loan applicants, their personal information, and the decision tree was eventually made.

This is a classic example of the type of problem tackled in the branch of AI known as *machine learning* (a sub-field in expert systems research). The challenge is to produce an algorithm that will organize a complex database in terms of some attribute we are particularly interested in such as an applicant's loan-worthiness, in the example we are considering. The organization takes the form of a decision tree, which will determine whether or not the attribute holds in a given case (i.e. whether or not the applicant is loan-worthy).

In the case of the mortgage loan decision tree the target attribute is labeled as *Loan*. All the branches of the decision tree must end in terminal nodes that have a value for the target attribute (i.e. they must say Yes or No). The decision tree is constructed by classifying the database in terms of other features such as *Good credit?* or *Earns more than \$70K?*. Once the decision tree has been constructed, it can then be used to decide whether some individual (e.g. a potential mortgage applicant) has the target attribute or not (i.e. is approved for the loan or not).

In the next section we will look in some detail at how an influential machine learning algorithm works. But first let me make explicit the connection with the physical symbol system hypothesis. As we saw in section 6.1, the physical symbol system hypothesis involves four basic claims:

¹ Symbols are physical patterns.

² Symbols can be combined to form complex symbol structures.

- **Summaries, checklists, and further reading.** These can be found at the end of each chapter. The summary provides a short overview of the chapter. The checklist allows students to review the key points of the chapter and also serves as a reference point for instructors. Suggestions of additional books and articles are provided to guide students' further reading on the topics covered in the chapter.



CHAPTER THIRTEEN

New horizons: Dynamical systems and situated cognition

OVERVIEW 403

13.1 Cognitive science and dynamical systems 404	13.3 Situated cognition and biobototics 420
What are dynamical systems? 405	The need for building a situated agent 421
The dynamical systems hypothesis: Cognitive science without representations? 406	Situated cognition and knowledge representation 421
13.2 Applying dynamical systems: Two case studies 412	Biobotic insects and morphological computation 424
Two ways of thinking about motor control 412	13.4 From subsumption architectures to multi-agent robotics 430
Dynamical systems and the A-not-B error 413	Subsumption architectures: The example of AIBO 430
Adapting the dynamical systems approach 419	Behavior-based robotics: TOTO 415
	Multi-agent programming: The Nerd Hero 438

Overview

Throughout this book we have been working through some of the basic consequences of a single principle. This is the principle that cognition is information processing. It is in many ways the most important framework assumption of cognitive science. The historical overview in Part I explored how researchers from a number of different disciplines converged on the information-processing model of cognition in the middle of the twentieth century. In Part II we looked at different ways of thinking about information processing – the physical symbol system hypothesis and the neural networks model. Despite their very significant differences, the physical symbol system and neural network approaches share a fundamental commitment to the idea that

403



Course Website

A course website accompanies the book. It can be found at www.cambridge.org/bermudez3. This website contains

- a bank of test questions
- PowerPoint slides for each chapter, organized by section
- electronic versions of figures from the text
- review questions for each chapter that students can use to check their understanding and to review the material
- sample syllabi for courses of different lengths and different levels
- links to useful learning resources, videos, and experimental demonstrations
- links to online versions of relevant papers and online discussions for each chapter

Instructors can access a password-protected section of the website. This contains

- suggested solutions for the more challenging exercises and problems

The website is a work in progress. Students and instructors are welcome to contact me with suggestions, revisions, and comments. Contact details are on the website.

Cart (0) Wishlist

Academic
The future of publishing since 1584

Search for keyword, author, ISBN, etc.

Include historic titles Welcome, Charles Sign Out

Subjects Textbooks Reference Authors News Conferences Blogs Contact Us

Home / Academic / Textbooks / Bermudez: Cognitive Science, 2nd edition

0

Bermudez: Cognitive Science, 2nd edition

[Home](#)

For students:
Experimental demonstrations
Videos
Testbanks
Useful links

For instructors:
PPT slides
Class discussion topics
Sample syllabi
Figures

José Luis Bermúdez
Cognitive Science **Second Edition**

Welcome to the resources site for 'Cognitive Science, 2nd edition'.

This highly successful textbook unifies the interdisciplinary streams of cognitive science into a single narrative. With clarity and breadth, it leads students to access, understand and apply the techniques and theories of the cognitive scientist's "toolkit" – the vast range of techniques and tools that cognitive scientists use to study the mind.

This Second Edition features a new chapter on consciousness, and additional online learning materials.



ACKNOWLEDGMENTS FOR THE FIRST EDITION

Many friends and colleagues associated with the Philosophy–Neuroscience–Psychology program at Washington University in St. Louis have commented on sections of this book. I would particularly like to thank Maurizio Corbetta, Frederick Eberhardt, David Kaplan, Clare Palmer, Gualtiero Piccinnini, Marc Raichle, Philip Robbins, David Van Essen, and Jeff Zacks. Josef Perner kindly read a draft of [Chapter 12](#).

I have benefited from the comments of many referees while working on this project. Most remain anonymous, but some have revealed their identity. My thanks to Kirsten Andrews, Gary Bradshaw, Rob Goldstone, Paul Humphreys, and Michael Spivey.

Drafts of this textbook have been used four times to teach PNP 200 Introduction to Cognitive Science here at Washington University in St. Louis – twice by me and once each by David Kaplan and Jake Beck. Feedback from students both inside and outside the classroom was extremely useful. I hope that other instructors who use this text have equally motivated and enthusiastic classes. I would like to record my thanks to the teaching assistants who have worked with me on this course: Juan Montaña, Tim Oakberg, Adam Shriver, and Isaac Wiegman. And also to Kimberly Mount, the PNP administrative assistant, whose help with the figures and preparing the manuscript is greatly appreciated.

A number of students from my Spring 2009 PNP 200 class contributed to the glossary. It was a pleasure to work with Olivia Frosch, Katie Lewis, Juan Manfredi, Eric Potter, and Katie Sadow.

Work on this book has been made much easier by the efforts of the Psychology textbook team at Cambridge University Press – Raihanah Begum, Catherine Flack, Hetty Reid, Sarah Wightman, and Rachel Willsher (as well as to Andy Peart, who signed this book up but has since moved on). They have been very patient and very helpful. My thanks also to Anna Oxbury for her editing and to Liz Davey for coordinating the production process.



ACKNOWLEDGMENTS FOR THE SECOND EDITION

I am very grateful to my colleagues in the Office of the Dean at Texas A&M University, particularly my administrative assistant Connie Davenport, for helping me to carve out time to work on the second edition of the textbook. T. J. Kasperbauer has been an excellent research assistant, providing numerous improvements to the text and supporting resources and helping me greatly with his deep knowledge of cognitive science. It has been a pleasure to work once again with Hetty Marx and Carrie Parkinson at Cambridge University Press. I particularly appreciate their work gathering feedback on the first edition. Thanks again to Anna Oxbury for her copyediting skills.



ACKNOWLEDGMENTS FOR THE THIRD EDITION

Part of the work on the third edition was carried out during a period of leave in the 2018–19 academic year. I am grateful to the Department of Philosophy at Texas A&M University for allowing me to take a year from teaching and administrative duties. At CUP, I worked initially with Janka Romero and then subsequently with Stephen Acerra and Lisa Pinto, all of whom were very supportive and worked hard to get me helpful feedback on earlier editions. I am particularly grateful to the many reviewers who gave detailed comments and suggestions for improvement. Finally, I'd like to record my thanks to Dong An for her help with the online resources.





Introduction

The Challenge of Cognitive Science

OVERVIEW 3

0.1 Cognitive Science: An Interdisciplinary Endeavor 3

0.2 Levels of Explanation: The Contrast between Psychology and Neuroscience 5

How Psychology Is Organized 6

How Neuroscience Is Organized 7

0.3 The Challenge of Cognitive Science 10

Three Dimensions of Variation 10

The Space of Cognitive Science 10



Overview

Cognitive science draws upon the tools and techniques of many different disciplines, including psychology, philosophy, linguistics, computer science, neuroscience, mathematical logic . . . It is a fundamentally *interdisciplinary activity*. This basic fact raises important and fundamental questions. What do all these disciplines have in common? How can they all come together to form a distinctive area of inquiry?

The aim of this introduction is to give you a sense of the scope and range of cognitive science, setting the framework for more detailed study in subsequent chapters. We will explore the idea that the different disciplines in cognitive science each study different levels of organization in the mind and the nervous system. In particular, we will see how the brain can be studied at many different levels, from the level of the molecule upward. The introduction ends with a description (and illustration) of what I call the space of cognitive science.



0.1

Cognitive Science: An Interdisciplinary Endeavor

The hexagonal diagram in Figure 0.1 is one of the most famous images in cognitive science. It comes from the 1978 report on the state of the art in cognitive science commissioned by the Sloan Foundation and written by a group of leading scholars. The diagram is intended to illustrate the interdisciplinary nature of cognitive science. The lines on the diagram

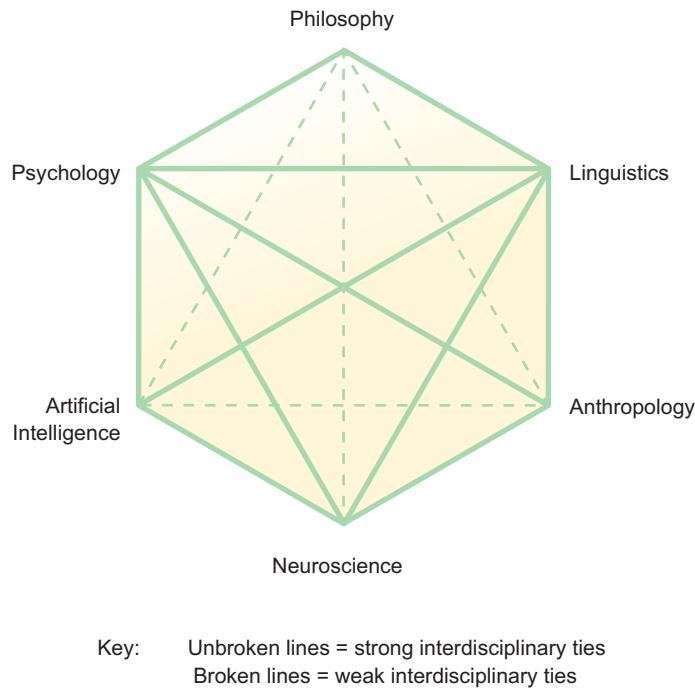


Figure 0.1 Connections among the cognitive sciences, as depicted in the Sloan Foundation's 1978 report. Unbroken lines indicate strong interdisciplinary links, while broken lines indicate weaker links. (Adapted from Gardner 1985)

indicate the academic disciplines that the authors saw as integral parts of cognitive science, together with the connections between disciplines particularly relevant to the study of mind and cognition.

For the authors of the Sloan report, cognitive science is an amalgamation of philosophy, psychology, linguistics, anthropology, neuroscience, and artificial intelligence. Each of the six disciplines brings with it different techniques, tools, and frameworks for thinking about the mind. Each of them studies the mind from different perspectives and at different levels. Whereas linguists, for example, develop abstract models of linguistic *competence* (the abstract structure of language), psychologists of language are interested in the mechanisms that make possible the *performance* of language users. Whereas neuroscientists study the details of how the brain works, computer scientists abstract away from those details to explore computer models and simulations of human cognitive abilities. Anthropologists are interested in the social dimensions of cognition, as well as how cognition varies across cultures. Philosophers, in contrast, are typically interested in very abstract models of how the mind is realized by the brain.

Some of the connections identified in the diagram were judged stronger than others. These are marked with a solid line. The weaker connections are marked with a broken line. At least one of the connections that was judged weak in 1978 has now become a



thriving subdiscipline in its own right. A group of philosophers impressed by the potential for fruitful dialog between philosophy and neuroscience have taken to calling themselves neurophilosophers, after the title of a very influential book by Patricia Churchland published in 1986.

Miller's own account of how the Sloan report was written is both disarming and telling. "The committee met once, in Kansas City. It quickly became apparent that everyone knew his own field and had heard of two or three interesting findings in other fields. After hours of discussion, experts in discipline X grew unwilling to make any judgments about discipline Y, and so forth. In the end, they did what they were competent to do: each summarized his or her own field and the editors – Samuel Jay Keyser, Edward Walker and myself – patched together a report" (Miller 2003: 143). This may be how reports get written, but it is not a very good model for an interdisciplinary enterprise such as cognitive science.

In fact, the hexagon as a whole is not a very good model for cognitive science. Even if we take seriously the lines that mark connections between the disciplines of cognitive science, the hexagon gives no sense of a unified intellectual enterprise. It gives no sense, that is, of something that is more than a composite of "traditional" disciplines such as philosophy and psychology. There are many different schools of philosophy and many different specializations within psychology, but there are certain things that bind together philosophers as a group and psychologists as a group, irrespective of their school and specialization. For philosophers (particularly in the so-called *analytic* tradition, the tradition most relevant to cognitive science), the unity of their discipline comes from certain problems that are standardly accepted as philosophical, together with a commitment to rigorous argument and analysis. The unity of psychology comes, in contrast, from a shared set of experimental techniques and paradigms. Is there anything that can provide a similar unity for cognitive science?

One of the main claims of this textbook is that cognitive science is indeed a unified enterprise. It has its own distinctive problems. Its own distinctive techniques, And its own distinctive explanatory frameworks. We will be studying all of these in this book. First, though, we need to get a better picture of the range and scope of the enterprise. In the rest of this introduction I'll use psychology and neuroscience as examples to give you a sense of the overall space of cognitive science.



0.2

Levels of Explanation: The Contrast between Psychology and Neuroscience

Neuroscience occupies one pole of the Sloan report's hexagonal figure and it was not viewed as very central to cognitive science by the authors of the report. The report was written before the "turn to the brain" that we will look at in [Chapter 3](#), and its focus reflected the contemporary focus on computer science, psychology, and linguistics as the core disciplines of cognitive science. Moreover, the authors of the report treated neuroscience as a unitary discipline, on a par with anthropology, psychology, and other more traditional academic disciplines. The explosion of research into what became known as cognitive neuroscience has since corrected both of these assumptions.



Most cognitive scientists place the study of the brain firmly at the heart of cognitive science. And it is becoming very clear that neuroscience is itself a massively interdisciplinary field.

How Psychology Is Organized

One way of thinking about what distinguishes neuroscience from, say, psychology is through the idea of levels. I am talking here about what is sometimes called scientific psychology (psychology as it is taught and studied in university departments), as opposed, for example, to humanistic psychology, self-help psychology, and much of what is routinely classified as psychology in bookstores. But even narrowing it down like this, there are many different subfields of psychology.

A quick look at the courses on offer in any reputable psychology department will find courses in cognitive psychology, social psychology, abnormal psychology, personality psychology, psychology of language, and so on. It is normal for research psychologists to specialize in at most one or two of these fields. Nonetheless, most psychologists think that psychology is a single academic discipline. This is partly because there is a continuity of methodology across the different specializations and subfields. Students in psychology are typically required to take a course in research methods. Such courses cover basic principles of experimental design, hypothesis formation and testing, and data analysis that are common to all branches of psychology.

Equally important, however, is the fact that many of these branches of psychology operate at the same level. The data from which they begin are data about cognitive performance and behavior at the level of the whole organism (I am talking about the whole organism to make clear that these ideas extend to nonhuman organisms, as studied in comparative psychology).

The basic *explananda* (the things that are to be explained) in psychology are people's psychological capacities, which includes both cognitive and emotional capacities. The organization of psychology into different subfields is a function of the fact that there are many different types of cognitive and emotional capacities.

Within cognitive psychology, for example, what psychologists are trying to explain are the organism's capacities for perception, memory, attention, and so on. Controlled experiments and correlational studies are used to delimit and describe those capacities, so that psychologists know exactly what it is that needs to be explained.

Likewise, social psychologists study the capacities involved in social understanding and social interactions. They are interested, for example, in social influences on behavior, on how we respond to social cues, and on how our thoughts and feelings are influenced by the presence of others. Personality psychologists study the traits and patterns of behavior that go to make up what we think of as a person's character. And so on.

If we were to map out some of the principal subfields in scientific psychology it would look something like [Figure 0.2](#). The diagram is intended to show that the different sub-branches all study different aspects of mind and behavior at the level of the organism.

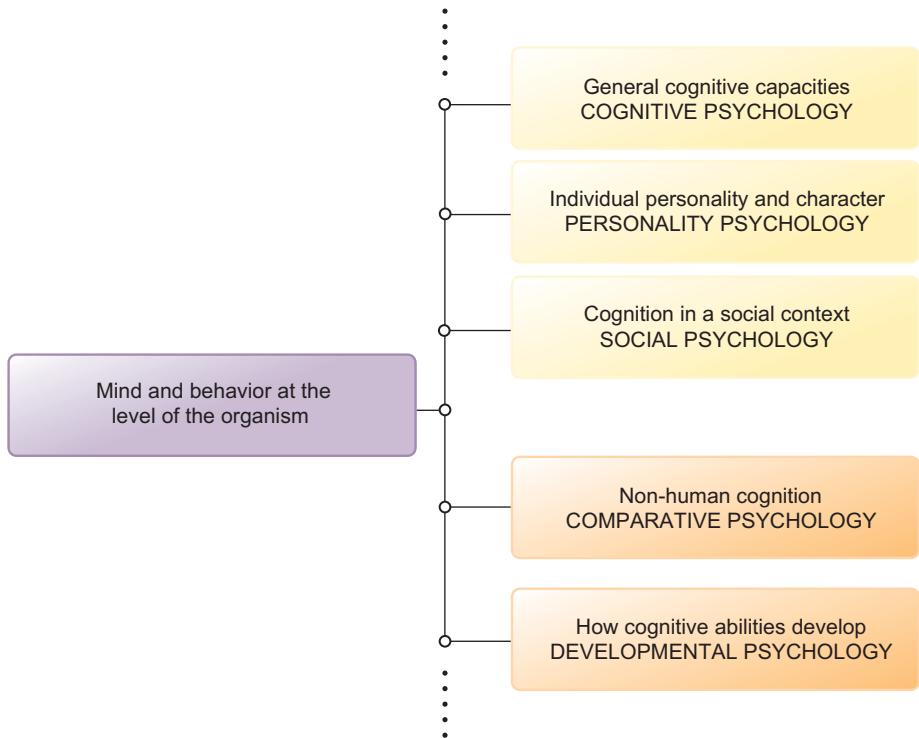


Figure 0.2 Some of the principal branches of scientific psychology.

How Neuroscience Is Organized

Things are very different in neuroscience. There are many branches of neuroscience, but they are not related in the same way. The organization of neuroscience into branches closely follows the different levels of organization in the brain and the central nervous system. These levels of organization are illustrated in [Figure 0.3](#), drawn from Gordon Shepherd's 1994 textbook *Neurobiology*.

You may have come across references to areas in the brain such as the primary visual cortex or the hippocampus, for example. And you may have encountered talk of neural pathways connecting different areas in the brain. Located at levels A and B in Shepherd's diagram, these are the highest levels of neural organization, corresponding most closely to cognitive activities that we all perform. The primary visual cortex, for example, is responsible for coding the basic features of visual information coming from the retina. It is sensitive to orientation, motion, speed, direction, and so on. The hippocampus, in contrast, is thought to be responsible for key aspects of memory.

Activity at this top level of organization is the result of activity at lower levels of organization. In Shepherd's diagram this takes us to levels C and E – the level of centers, local circuits, and microcircuits. Somehow the collective activity of populations of neurons codes certain types of information about objects in a way that organizes and coordinates

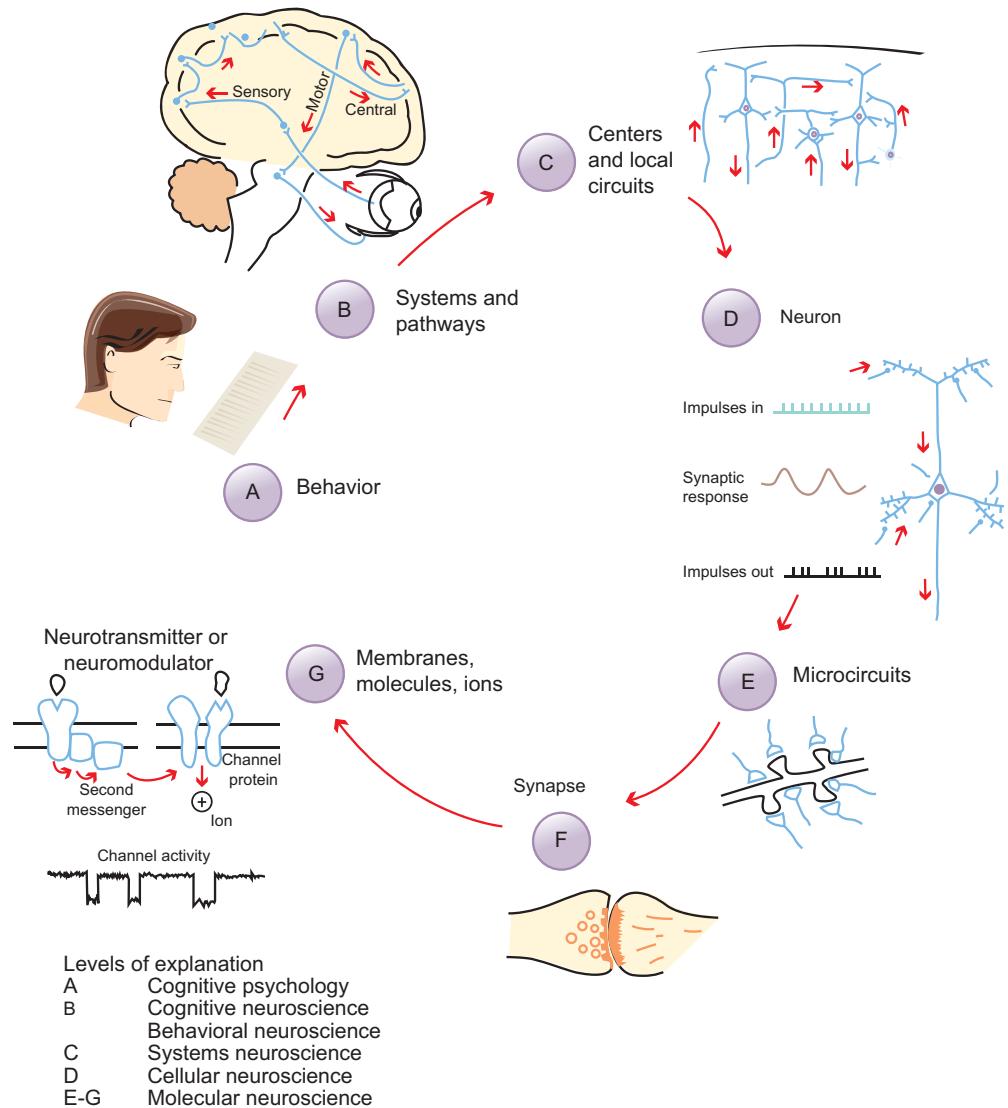


Figure 0.3 Levels of organization and levels of explanation in the nervous system. (Adapted from Shepherd 1994)

the information carried by individual neurons. These populations of neurons are the local circuits in Shepherd's diagram.

What happens in populations of neurons is ultimately determined by the behavior of individual neurons. But neurons are not the most basic level of organization in the nervous system. In order to understand how neurons work we need to understand how they communicate. This brings us to Shepherd's level F, because neurons communicate across synapses. Most synapses are chemical, but some are electrical. The chemical synapses work through the transmission of neurochemicals (*neurotransmitters*). These neurotransmitters are activated by



the arrival of an electrical signal (the *action potential*). The propagation of neurotransmitters works the way it does because of the molecular properties of the synaptic membrane—properties that are ultimately genetically determined. With this we arrive at level G in Shepherd's diagram.

The point of this whistle-stop tour through the levels of organization in the brain is that the subfields of neuroscience map very closely onto the different levels of organization in the brain. At the top level we have cognitive neuroscience and behavioral neuroscience, which study the large-scale organization of the brain circuits deployed in high-level cognitive activities. These operate at what in discussing the subfields of psychology I termed the level of the whole organism. Systems neuroscience, in contrast, investigates the functioning of neural systems, such as the visual system. The bridge between the activity of neural systems and the activity of individual neurons is one of the central topics in computational neuroscience, while cellular and molecular neuroscience deal with the fundamental biological properties of neurons.

Different branches of neuroscience (and cognitive science in general) employ tools appropriate to the level of organization at which they are studying the brain. These tools and techniques vary in what neuroscientists call their temporal and spatial resolution. That is, they vary in the scale on which they give precise measurements (spatial resolution) and the time intervals to which they are sensitive (temporal resolution).

Some of the important variations are depicted in [Figure 0.4](#). We will explore the differences between these different tools and technologies in much more detail in later chapters (particularly [Chapter 9](#)).

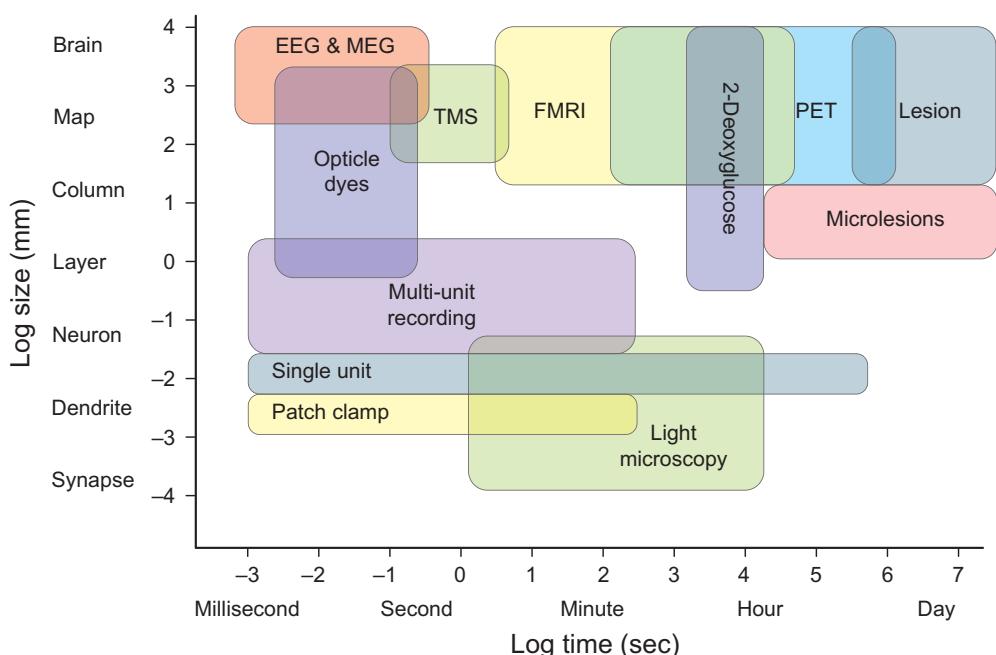


Figure 0.4 The spatial and temporal resolution of different tools and techniques in neuroscience. Time is on the x-axis and size is on the y-axis. (Adapted from Baars and Gage 2010)

 0.3

The Challenge of Cognitive Science

This section explores these basic ideas of levels of organization, levels of resolution, and levels of explanation further, to give a picture of what I call the space of cognitive science.



Three Dimensions of Variation

Cognitive science draws upon a large number of potentially relevant fields and subfields. Those fields and subfields differ from each other along three dimensions.

One dimension of variation is illustrated by the subfields of neuroscience. Neuroscience studies the brain at many different levels. These levels are organized into a vertical hierarchy that corresponds to the different levels of organization in the nervous system.

A second dimension of variation comes with the different techniques and tools that cognitive scientists can employ. As illustrated in [Figure 0.4](#), these tools vary both in spatial and in temporal resolution. Some tools, such as PET and fMRI, give accurate measurements at the level of individual brain areas. Others, such as microelectrode recording, give accurate measurements at the level of individual neurons (or small populations of neurons).

The third dimension of variation is exemplified by the different subfields of psychology. Most of psychology operates at Shepherd's level A. The different areas of psychology set out to explore, map, describe, and explain are the cognitive abilities making possible the myriad things that human beings do and say.



The Space of Cognitive Science

The different parts of cognitive science are distributed, therefore, across a three-dimensional space illustrated in [Figure 0.5](#).

- The x -axis marks the different cognitive domains that are being studied
- The y -axis marks the different tools that might be employed (ordered roughly in terms of their degree of spatial resolution).
- The z -axis marks the different levels of organization at which cognition is studied.

This three-dimensional diagram is a more accurate representation of where cognitive science stands in the early years of the twenty-first century than the two-dimensional hexagon proposed by the authors of the Sloan report (although the hexagon may well have been an adequate picture of how things stood at the end of the 1970s).

A good way of thinking about cognitive science is as setting out to provide a unified account of cognition that draws upon and integrates the whole space. Cognitive science is more than just the sum of its parts. The aim of cognitive science as an intellectual enterprise is to provide a framework that makes explicit the common ground between all the different academic disciplines that study the mind and that shows how they are related to each other.

You can think of the analogy with physics. Many theoretical physicists think that the ultimate goal of physics is to provide a unified Theory of Everything. So too (on this way of thinking about cognitive science) is it the mission of cognitive science to provide a unified Theory of Cognition.

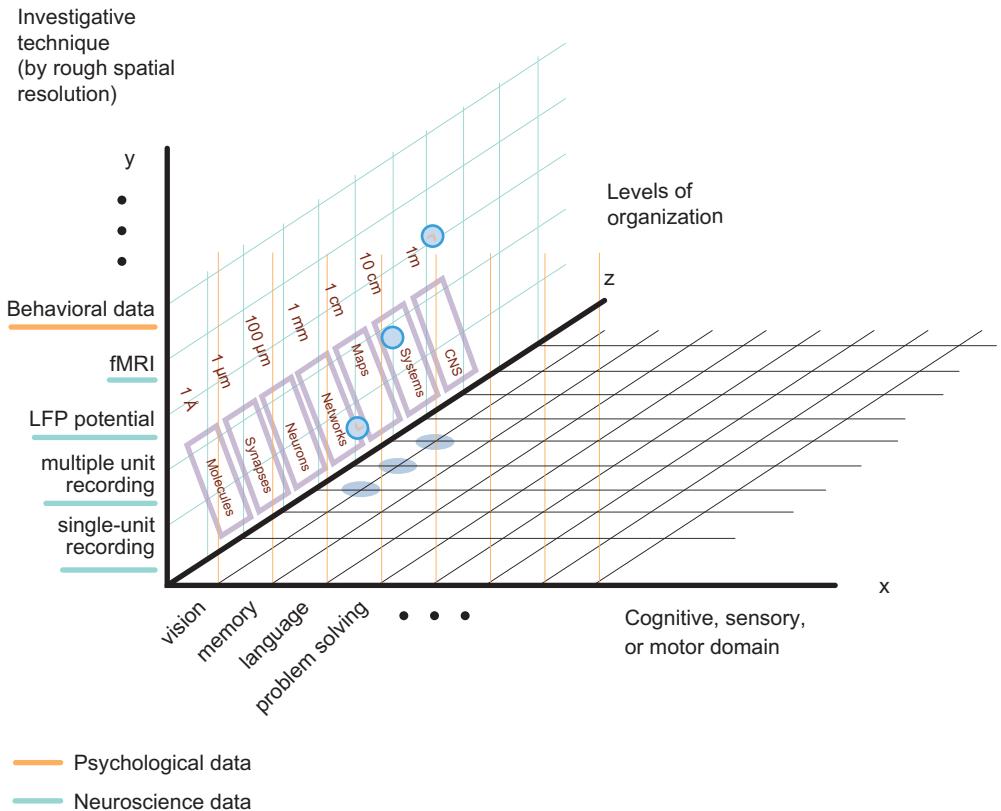


Figure 0.5 The “space” of contemporary cognitive science.

Parts II and III will explore the principal theories of cognition in cognitive science, and see how they can be applied to explain different aspects of cognition. First, though, we turn to an overview of some of the key historical landmarks in the emergence and subsequent development of cognitive science. That will occupy the three chapters of Part I. These chapters should put flesh on the bones of the general picture sketched out in this introduction.

Further Reading

Historical background on the Sloan report can be found in Gardner 1985 and Miller 2003 (available in the online resources). The report itself was never published. A very useful basic introduction to levels of organization and structure in the nervous system is chapter 2 of Churchland and Sejnowski 1992. For more detail, a classic neuroscience textbook is Kandel, Schwarz, and Jessell 2012. Stein and Stoodley 2006 and Purves et al. 2011 are alternatives. Craver 2007 discusses the interplay between different levels of explanation in the neuroscience of memory. Piccinini and Craver 2011 is a more general discussion; also see Bickle 2006 and Sullivan 2009.

PART I

HISTORICAL LANDMARKS









CHAPTER ONE

The Prehistory of Cognitive Science

OVERVIEW	15
1.1 The Reaction against Behaviorism in Psychology	16
Learning without Reinforcement: Tolman and Honzik, "Insight' in Rats" (1930)	17
Cognitive Maps in Rats? Tolman, Ritchie, and Kalish, "Studies in Spatial Learning" (1946)	20
Plans and Complex Behaviors: Lashley, "The Problem of Serial Order in Behavior" (1951)	21
1.2 The Theory of Computation and the Idea of an Algorithm	22
Algorithms and Turing Machines: Turing, "On Computable Numbers, with an Application to the Decision Problem" (1936–7)	23
1.3 Linguistics and the Formal Analysis of Language	25
The Structure of Language: Chomsky's <i>Syntactic Structures</i> (1957)	26
1.4 Information-Processing Models in Psychology	28
How Much Information Can We Handle? George Miller's "The Magical Number Seven, Plus or Minus Two" (1956)	29
The Flow of Information: Donald Broadbent's "The Role of Auditory Localization in Attention and Memory Span" (1954) and <i>Perception and Communication</i> (1958)	30
1.5 Connections and Points of Contact	32



Overview

In the late 1970s cognitive science became an established part of the intellectual landscape. At that time an academic field crystallized around a basic set of problems, techniques, and theoretical assumptions. These problems, techniques, and theoretical assumptions came from many different disciplines and areas. Many of them had been around for a fairly long time. What was new was the idea of putting them together as a way of studying the mind.

Cognitive science is at heart an interdisciplinary endeavor. In interdisciplinary research great innovations come about simply because people see how to combine things that are already out there but have never been put together before. A good way to understand cognitive science is to try to think your way back to how things might have looked to its early pioneers. They were exploring a landscape in which certain regions were well mapped and well understood, but where

there were no standard ways of getting from one region to another. An important part of what they did was to show how these different regions could be connected in order to create an interdisciplinary science of the mind.

In this chapter we go back to the 1930s, 1940s, and 1950s – to explore the *prehistory* of cognitive science. We will be looking at some of the basic ideas and currents of thought that, in retrospect, we can see as feeding into what came to be known as cognitive science. As we shall see in more detail later on in this book, *the guiding idea of cognitive science is that mental operations involve processing information*, and hence that we can study how the mind works by studying how information is processed. This basic idea of the mind as an information processor has a number of very specific roots, in areas that seem on the face of it to have little in common. The prehistory of cognitive science involves parallel, and largely independent, developments in psychology, linguistics, and mathematical logic. We will be looking at four of these developments:

- The reaction against behaviorism in psychology (Section 1.1)
- The idea of algorithmic computation in mathematical logic (Section 1.2)
- The emergence of linguistics as the formal analysis of language (Section 1.3)
- The emergence of information-processing models in psychology (Section 1.4)

In concentrating on these four developments we will be passing over other important influences, such as neuroscience and neuropsychology. This is because until quite recently the direct study of the brain had a relatively minor role to play in cognitive science.

Almost all cognitive scientists are convinced that in some fundamental sense the mind just is the brain, so that everything that happens in the mind is happening in the brain. Few, if any, cognitive scientists are *dualists*, who think that the mind and the brain are two separate and distinct things. But for a long time in the history of cognitive science it was widely held that we are better off studying the mind by abstracting away from the details of what is going on in the brain. This changed only with the emergence in the 1970s and 1980s of new technologies for studying neural activity and of new ways of modeling cognitive abilities – as we will see in Chapter 3.

1.1 The Reaction against Behaviorism in Psychology

Behaviorism was (and in some quarters still is) an influential movement in psychology. It takes many different forms, but they all share the basic assumption that psychologists should confine themselves to studying observable phenomena and measurable behavior. Behaviorists think that psychologists should avoid speculating about unobservable mental states, and instead focus on nonpsychological mechanisms linking particular stimuli with particular responses. These mechanisms are the product of conditioning. For examples of conditioning, think of Pavlov's dogs being conditioned to salivate at the sound of the bell, or the rewards/punishments that animal trainers use to encourage/discourage certain types of behavior.

For behaviorists, psychology is really the science of behavior. This approach to psychology leaves little room for cognitive science as the scientific study of cognition and the mind. Cognitive science could not even get started until behaviorism ceased to be the



dominant approach within psychology. Psychology's move from behaviorism was a lengthy and drawn-out process (and some would say that it has not yet been completed). We can appreciate some of the ideas that proved important for the later development of cognitive science by looking at three landmark papers. Each was an important statement of the idea that various types of behavior could not be explained in terms of stimulus-response mechanisms. Instead, psychologists need to think about organisms as storing and processing information about their environment, rather than as responding mechanically to reinforcers and stimuli. This idea of organisms as information processors is the single most fundamental idea of cognitive science.

Learning without Reinforcement: Tolman and Honzik, "Insight" in Rats" (1930)

Edward Tolman (1886–1959) was a behaviorist psychologist studying problem solving and learning in rats (among other things). As with most psychologists of the time, he started off with two standard behaviorist assumptions about learning. The first assumption is that all learning is the result of *conditioning*. The second assumption is that conditioning depends upon processes of *association* and *reinforcement*.

We can understand these two assumptions by thinking about a rat in what is known as a Skinner box, after the celebrated behaviorist B. F. Skinner. A typical Skinner box is illustrated in [Figure 1.1](#). The rat receives a reward each time it behaves in a particular way (pressing a lever, for example, or pushing a button). The reward *reinforces* the behavior. This means that the association between the behavior and the reward is strengthened and the rat's performing the behavior again becomes more likely. The rat becomes *conditioned* to perform the behavior.

The basic idea of behaviorism is that all learning is either reinforcement learning of this general type, or the even simpler form of associative learning often called classical conditioning.

In classical conditioning what is strengthened is the association between a *conditioned stimulus* (such as the typically neutral sound of a bell ringing) and an *unconditioned stimulus* (such as the presentation of food). The unconditioned stimulus is *not* neutral for the organism and typically provokes a behavioral response, such as salivation. What happens during classical conditioning is that the strengthening of the association between conditioned stimulus and unconditioned stimulus eventually leads the organism to produce the unconditioned response to the conditioned stimulus alone, without the presence of the unconditioned stimulus. The most famous example of classical conditioning is Pavlov's dogs, who were conditioned to salivate to the sound of a bell by the simple technique of using the bell to signal the arrival of food.

So, it is a basic principle of behaviorism that all learning, whether by rats or by human beings, takes place through processes of reinforcement and conditioning. What the studies reported by Tolman and Honzik in [1930](#) seemed to show, however, is that this is not true even for rats.

Tolman and Honzik were interested in how rats learned to navigate mazes. They ran three groups of rats through a maze of the type illustrated in [Figure 1.2](#). The first group

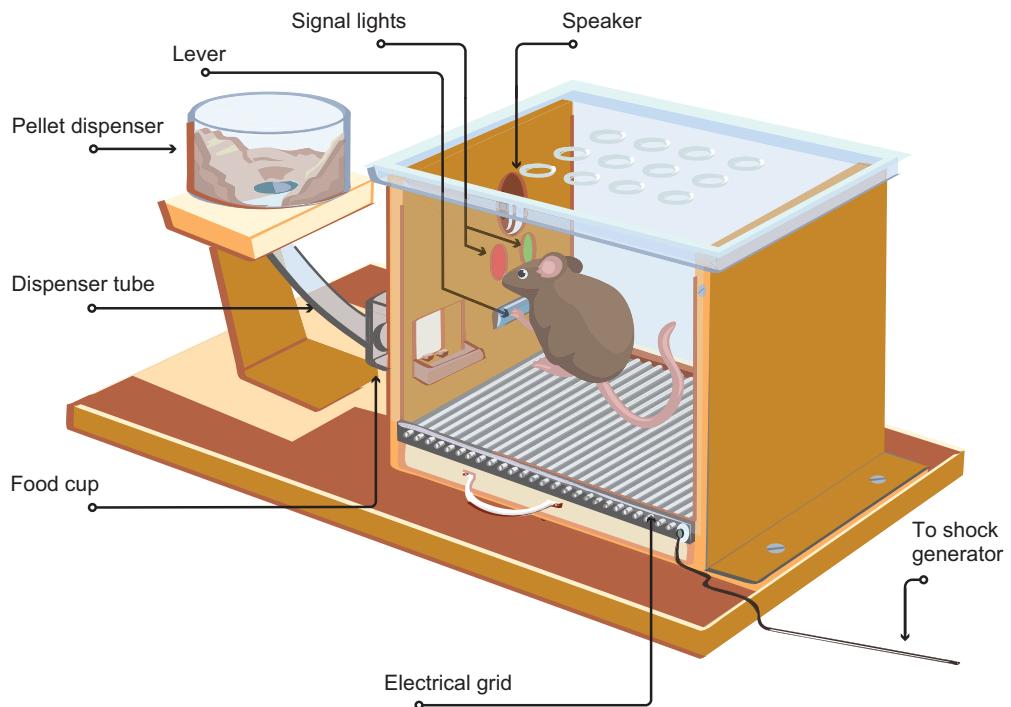


Figure 1.1 A rat in a Skinner box. The rat has a response lever controlling the delivery of food, as well as devices allowing different types of stimuli to be produced.

received a reward each time they successfully ran the maze. The second group never received a reward. The third group was unrewarded for the first ten days and then began to be rewarded.

As behaviorism predicted, the rewarded rats quickly learned to run the maze, while both groups of unrewarded rats simply wandered around aimlessly. The striking fact, however, was that when the third group of rats started to receive rewards they learned to run the maze far more quickly than the first group had.

Tolman and Honzik argued that the rats must have been learning about the layout of the maze during the period when they were not being rewarded. This type of *latent learning* seemed to show that reinforcement was not necessary for learning, and that the rats must have been picking up and storing information about the layout of the maze when they were wandering around it, even though there was no reward and hence no reinforcement. They were later able to use this information to navigate the maze.



Exercise 1.1 Explain in your own words why latent learning seems to be incompatible with the two basic assumptions of behaviorism.

Suppose, then, that organisms are capable of latent learning – that they can store information for later use without any process of reinforcement. One important follow-up

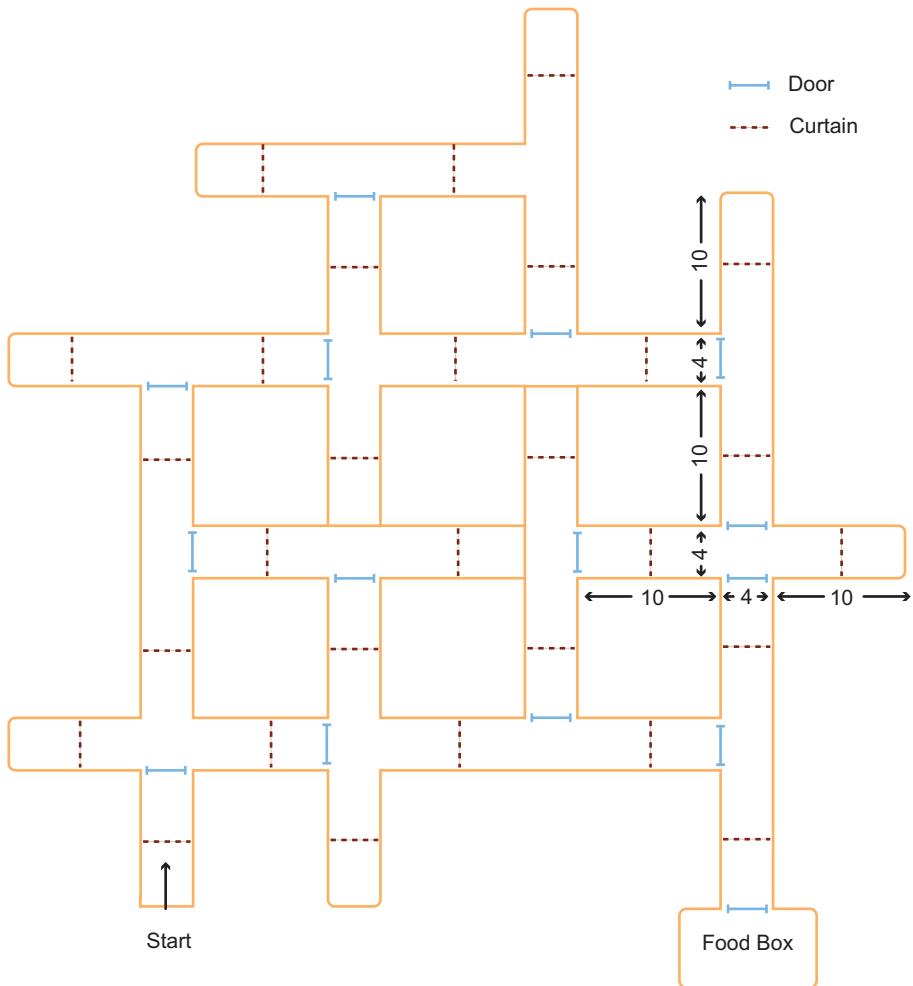


Figure 1.2 A fourteen-unit T-Alley maze (measurements in inches). Note the blocked passages and dead ends. (Adapted from Elliott 1928)

question is: What sort of information is being stored? In particular, are the rats storing information about the spatial layout of the maze? Or are they simply “remembering” the sequences of movements (responses) that they made while wandering around the maze? And so, when the rats in the latent-learning experiments start running the maze successfully, are they simply repeating their earlier sequences of movements, or are they using their “knowledge” of how the different parts of the maze fit together?

Tolman and his students and collaborators designed many experiments during the 1930s and 1940s to try to decide between *place learning* and *response learning* accounts of how rats learn to run a maze. Some of these experiments were reported in a famous article in 1946.