



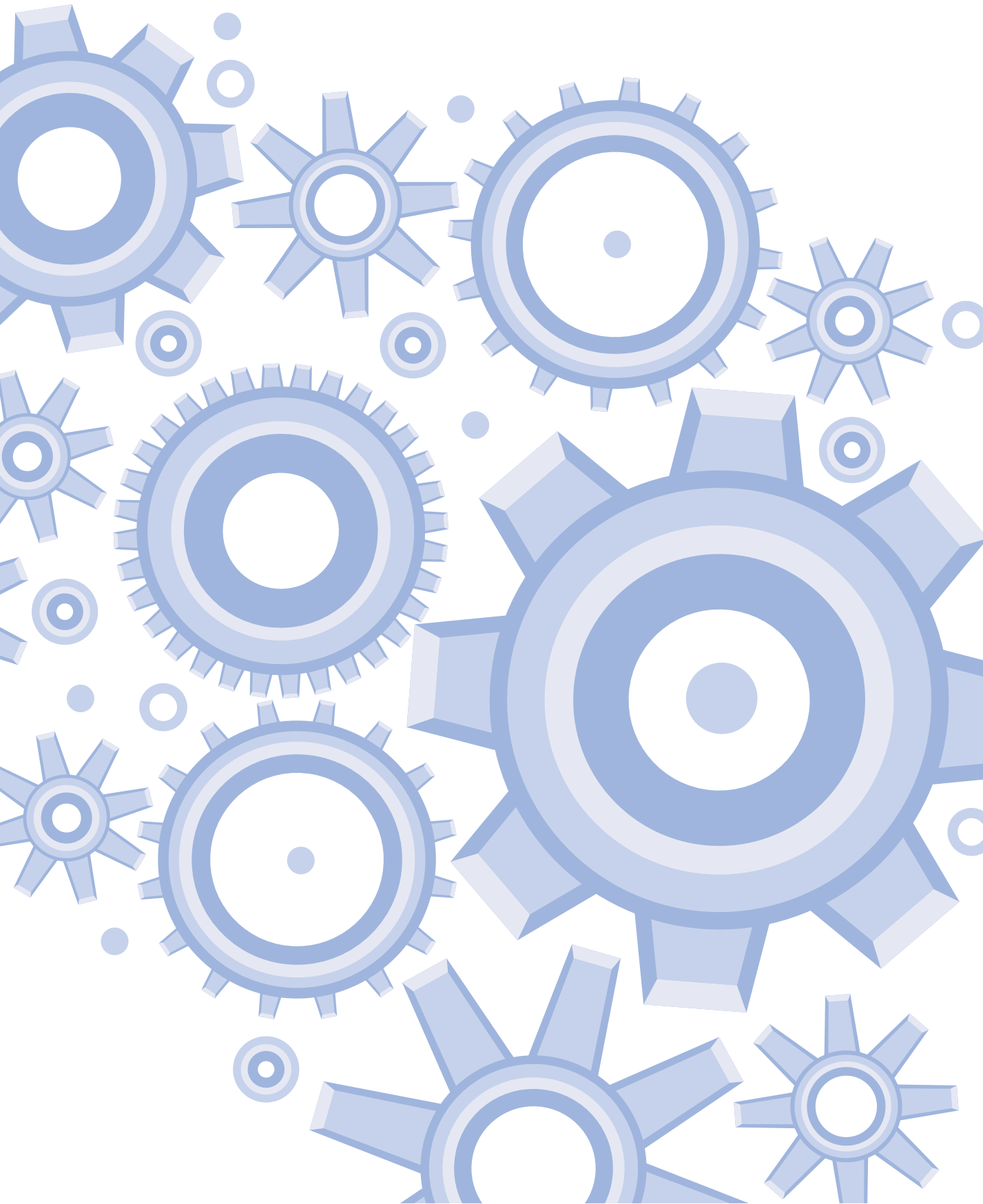
## COGNITIVE SCIENCE

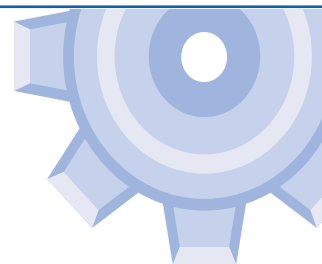
### Second edition

*Cognitive Science* combines the interdisciplinary streams of cognitive science into a unified narrative in an all-encompassing introduction to the field. This text presents cognitive science as a discipline in its own right, and teaches students to apply the techniques and theories of the cognitive scientist's "toolkit" – the vast range of methods and tools that cognitive scientists use to study the mind. Thematically organized, rather than by separate disciplines, *Cognitive Science* underscores the problems and solutions of cognitive science, rather than those of the subjects that contribute to it – psychology, neuroscience, linguistics, etc. 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 updated and enhanced to incorporate new studies and key experiments since the first edition. A new chapter on consciousness has been added.

JOSÉ LUIS BERMÚDEZ is Dean of the College of Liberal Arts and Professor of Philosophy at Texas A&M University. He has been involved in teaching and research in cognitive science for over twenty years, and is very much involved in bringing an interdisciplinary focus to cognitive science through involvement with conference organization and journals. His 100+ publications include the textbook *Philosophy of Psychology: A Contemporary Introduction* (2005) and a companion collection of readings, *Philosophy of Psychology: Contemporary Readings* (2007). He has authored the monographs *The Paradox of Self-Consciousness* (1998), *Thinking without Words* (2003), and *Decision Theory and Rationality* (2009) in addition to editing a number of collections including *The Body and the Self* (1995), *Reason and Nature* (2002), and *Thought, Reference, and Experience* (2005).

Cambridge University Press  
978-1-107-05162-1 — Cognitive Science  
2nd Edition  
Frontmatter  
[More Information](#)



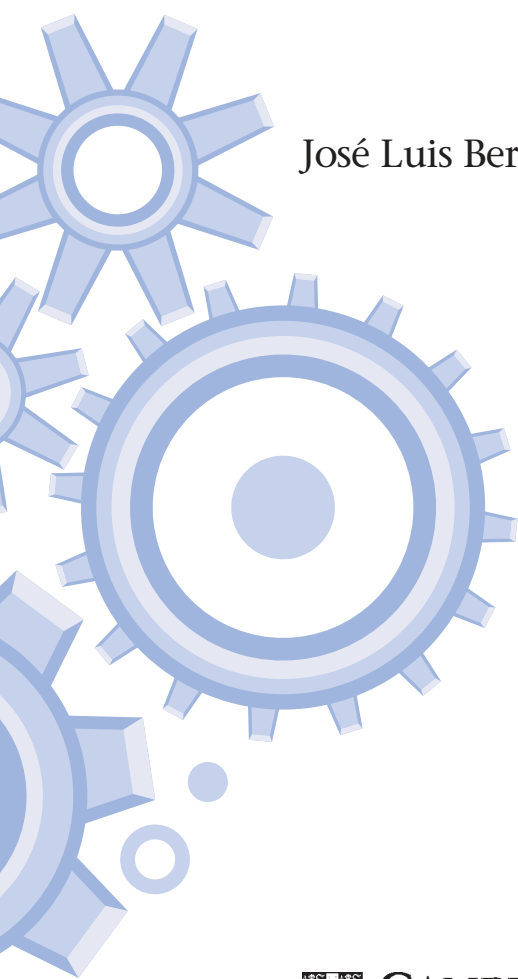


# COGNITIVE SCIENCE

An Introduction to the Science of the Mind

Second Edition

José Luis Bermúdez



CAMBRIDGE  
UNIVERSITY PRESS

**CAMBRIDGE**  
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

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](http://www.cambridge.org)

Information on this title: [www.cambridge.org/9781107653351](http://www.cambridge.org/9781107653351)

© José Luis Bermúdez 2014

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 2010

Second edition 2014

5th printing 2017

Printed in the United Kingdom by Bell and Bain Ltd, Glasgow

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

*Library of Congress Cataloguing in Publication data*

Bermúdez, José Luis.

Cognitive science : an introduction to the science of the mind / José Luis Bermúdez. – Second edition.

pages cm

ISBN 978-1-107-05162-1 (Hardback) – ISBN 978-1-107-65335-1 (Paperback) 1. Cognition.

2. Cognitive science. I. Title.

BF311.B458 2014

153–dc23 2013047706

ISBN 978-1-107-05162-1 Hardback

ISBN 978-1-107-65335-1 Paperback

Additional resources for this publication at [www.cambridge.org/bermudez2](http://www.cambridge.org/bermudez2)

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*   xxv  
*Preface*   xxvii  
*Acknowledgments for the first edition*   xxxiv  
*Acknowledgments for the second edition*   xxxv

**PART I   HISTORICAL LANDMARKS   2**  
*Introduction to Part I*   3  
1   The prehistory of cognitive science   5  
2   The discipline matures: Three milestones   29  
3   The turn to the brain   59

**PART II   THE INTEGRATION CHALLENGE   84**  
*Introduction to Part II*   85  
4   Cognitive science and the integration challenge   87  
5   Tackling the integration challenge   113

**PART III   INFORMATION-PROCESSING MODELS OF THE MIND   138**  
*Introduction to Part III*   139  
6   Physical symbol systems and the language of thought   141  
7   Applying the symbolic paradigm   171  
8   Neural networks and distributed information processing   209  
9   Neural network models of cognitive processes   239

**PART IV   THE ORGANIZATION OF THE MIND   276**  
*Introduction to Part IV*   277  
10   How are cognitive systems organized?   279  
11   Strategies for brain mapping   315  
12   A case study: Exploring mindreading   353

**PART V NEW HORIZONS 400**  
*Introduction to Part V 401*  
**13** New horizons: Dynamical systems and situated cognition 403  
**14** The cognitive science of consciousness 445  
**15** Looking ahead: Challenges and applications 481

*Glossary 486*  
*Bibliography 495*  
*Index 514*



CONTENTS

*List of boxes*   xiv  
*List of figures*   xv  
*List of tables*   xxv  
*Preface*   xxvii  
*Acknowledgments for the first edition*   xxxiv  
*Acknowledgments for the second edition*   xxxv

PART I   HISTORICAL LANDMARKS   2

*Introduction to Part I*   3

1   The prehistory of cognitive science   5

1.1   The reaction against behaviorism in psychology   6

Learning without reinforcement: Tolman and Honzik, “‘Insight’ in rats” (1930)   7  
Cognitive maps in rats? Tolman, Ritchie, and Kalish, “Studies in spatial learning” (1946)   10

Plans and complex behaviors: Lashley, “The problem of serial order in behavior” (1951)   12

1.2   The theory of computation and the idea of an algorithm   13

Algorithms and Turing machines: Turing, “On computable numbers, with an application to the Decision Problem” (1936-7)   13

1.3   Linguistics and the formal analysis of language   16

The structure of language: Chomsky’s *Syntactic Structures* (1957)   16

1.4   Information-processing models in psychology   19

How much information can we handle? George Miller’s “The magical number seven, plus or minus two” (1956)   19

The flow of information: Donald Broadbent’s “The role of auditory localization in attention and memory span” (1954) and *Perception and Communication* (1958)   21

1.5   Connections and points of contact   23

2   The discipline matures: Three milestones   29

2.1   Language and micro-worlds   30

Natural language processing: Winograd, *Understanding Natural Language* (1972)   31  
SHRDLU in action   33

2.2	How do mental images represent?	39
	Mental rotation: Shepard and Metzler, “Mental rotation of three-dimensional objects” (1971)	40
	Information processing in mental imagery	43
2.3	An interdisciplinary model of vision	46
	Levels of explanation: Marr’s <i>Vision</i> (1982)	46
	Applying top-down analysis to the visual system	48
3	The turn to the brain	59
3.1	Cognitive systems as functional systems	60
3.2	The anatomy of the brain and the primary visual pathway	62
	The two visual systems hypothesis: Ungerleider and Mishkin, “Two cortical visual systems” (1982)	65
3.3	Extending computational modeling to the brain	70
	A new set of algorithms: Rumelhart, McClelland, and the PDP Research Group, <i>Parallel Distributed Processing: Explorations in the Microstructure of Cognition</i> (1986)	72
	Pattern recognition in neural networks: Gorman and Sejnowski’s mine/rock detector	74
3.4	Mapping the stages of lexical processing	76
	Functional neuroimaging	77
	Petersen <i>et al.</i> , “Positron emission tomographic studies of the cortical anatomy of single-word processing” (1988)	78
PART II	THE INTEGRATION CHALLENGE	84
	Introduction to Part II	85
4	Cognitive science and the integration challenge	87
4.1	Cognitive science: An interdisciplinary endeavor	88
4.2	Levels of explanation: The contrast between psychology and neuroscience	91
	How psychology is organized	91
	How neuroscience is organized	93
4.3	The integration challenge	95
	How the fields and sub-fields vary	96
	The space of cognitive science	97
4.4	Local integration I: Evolutionary psychology and the psychology of reasoning	99
	Conditional reasoning	100
	The reasoning behind cooperation and cheating: The prisoner’s dilemma	102
4.5	Local integration II: Neural activity and the BOLD signal	105



5	Tackling the integration challenge	113
5.1	Intertheoretic reduction and the integration challenge	114
	What is intertheoretic reduction?	115
	The prospects for intertheoretic reduction in cognitive science	116
5.2	Marr’s tri-level hypothesis and the integration challenge	122
	Problems with the tri-level hypothesis as a blueprint for cognitive science	126
5.3	Models of mental architecture	129
	Modeling information processing	130
	Modeling the overall structure of the mind	131
PART III	INFORMATION-PROCESSING MODELS OF THE MIND	138
	Introduction to Part III	139
6	Physical symbol systems and the language of thought	141
6.1	The physical symbol system hypothesis	142
	Symbols and symbol systems	144
	Solving problems by transforming symbol structures	144
	Intelligent action and the physical symbol system	150
6.2	From physical symbol systems to the language of thought	151
	Intentional realism and causation by content	153
	The computer model of the mind and the relation between syntax and semantics	155
	Putting the pieces together: Syntax and the language of thought	157
6.3	The Chinese room argument	160
	The Chinese room and the Turing test	162
	Responding to the Chinese room argument	163
	The symbol-grounding problem	165
7	Applying the symbolic paradigm	171
7.1	Expert systems, machine learning, and the heuristic search hypothesis	172
	Expert systems and decision trees	173
	Machine learning and the physical symbol system hypothesis	175
7.2	ID3: An algorithm for machine learning	176
	From database to decision tree	177
	ID3 in action	181
	ID3 and the physical symbol system hypothesis	186
7.3	WHISPER: Predicting stability in a block world	188
	WHISPER: How it works	189
	WHISPER solving the chain reaction problem	191
	WHISPER: What we learn	195

	<b>7.4 Putting it all together: SHAKEY the robot</b>	196
	SHAKEY's software I: Low-level activities and intermediate-level actions	197
	SHAKEY's software II: Logic programming in STRIPS and PLANEX	201
<b>8</b>	<b>Neural networks and distributed information processing</b>	209
	<b>8.1 Neurally inspired models of information processing</b>	210
	Neurons and network units	212
	<b>8.2 Single-layer networks and Boolean functions</b>	216
	Learning in single-layer networks: The perceptron convergence rule	220
	Linear separability and the limits of perceptron convergence	223
	<b>8.3 Multilayer networks</b>	227
	The backpropagation algorithm	229
	How biologically plausible are neural networks?	230
	<b>8.4 Information processing in neural networks: Key features</b>	232
	Distributed representations	232
	No clear distinction between information storage and information processing	233
	The ability to learn from "experience"	235
<b>9</b>	<b>Neural network models of cognitive processes</b>	239
	<b>9.1 Language and rules: The challenge for information-processing models</b>	240
	What is it to understand a language?	241
	Language learning and the language of thought: Fodor's argument	243
	<b>9.2 Language learning in neural networks</b>	245
	The challenge of tense learning	246
	Neural network models of tense learning	249
	<b>9.3 Object permanence and physical reasoning in infancy</b>	254
	Infant cognition and the dishabituation paradigm	255
	How should the dishabituation experiments be interpreted?	260
	<b>9.4 Neural network models of children's physical reasoning</b>	261
	Modeling object permanence	263
	Modeling the balance beam problem	266
	<b>9.5 Conclusion: The question of levels</b>	269
<b>PART IV</b>	<b>THE ORGANIZATION OF THE MIND</b>	276
	<i>Introduction to Part IV</i>	277
<b>10</b>	<b>How are cognitive systems organized?</b>	279
	<b>10.1 Architectures for intelligent agents</b>	280
	Three agent architectures	281

- 10.2 Fodor on the modularity of mind** 285
    - Characteristics of modular processing 288
    - Central processing 290
    - Modularity and cognitive science 291
  - 10.3 The massive modularity hypothesis** 294
    - From reasoning experiments to Darwinian modules 295
    - The argument from error 298
    - The argument from statistics and learning 298
    - Evaluating the arguments for massive modularity 301
  - 10.4 Hybrid architectures** 305
    - The ACT-R/PM architecture 306
    - ACT-R/PM as a hybrid architecture 308
- 11 Strategies for brain mapping** 315
  - 11.1 Structure and function in the brain** 316
    - Exploring anatomical connectivity 318
  - 11.2 Studying cognitive functioning: Techniques from neuroscience** 324
    - Mapping the brain's electrical activity: EEG and MEG 325
    - Mapping the brain's blood flow and blood oxygen levels: PET and fMRI 329
  - 11.3 Combining resources I: The locus of selection problem** 330
    - Combining ERPs and single-unit recordings 332
  - 11.4 Combining resources II: Networks for attention** 337
    - Two hypotheses about visuospatial attention 339
  - 11.5 From data to maps: Problems and pitfalls** 343
    - From blood flow to cognition? 343
    - Noise in the system? 344
    - Functional connectivity vs. effective connectivity 345
- 12 A case study: Exploring mindreading** 353
  - 12.1 Pretend play and metarepresentation** 354
    - The significance of pretend play 355
    - Leslie on pretend play and metarepresentation 356
    - The link to mindreading 360
  - 12.2 Metarepresentation, autism, and theory of mind** 361
    - Using the false belief task to study mindreading 362
    - Interpreting the results 364
    - Implicit and explicit understanding of false belief 366

	<b>12.3 The mindreading system</b>	368
	First steps in mindreading	369
	From dyadic to triadic interactions: Joint visual attention	371
	TESS and TOMM	372
	<b>12.4 Understanding false belief</b>	374
	The selection processor hypothesis	374
	An alternative model of theory of mind development	376
	<b>12.5 Mindreading as simulation</b>	381
	Standard simulationism	382
	Radical simulationism	384
	<b>12.6 The cognitive neuroscience of mindreading</b>	385
	Neuroimaging evidence for a dedicated theory of mind system?	386
	Neuroscientific evidence for simulation in low-level mindreading?	390
	Neuroscientific evidence for simulation in high-level mindreading?	394
<b>PART V</b>	<b>NEW HORIZONS</b>	400
	<i>Introduction to Part V</i>	401
<b>13</b>	<b>New horizons: Dynamical systems and situated cognition</b>	403
	<b>13.1 Cognitive science and dynamical systems</b>	404
	What are dynamical systems?	405
	The dynamical systems hypothesis: Cognitive science without representations?	406
	<b>13.2 Applying dynamical systems: Two examples from child development</b>	412
	Two ways of thinking about motor control	412
	Dynamical systems and the A-not-B error	414
	Assessing the dynamical systems approach	419
	<b>13.3 Situated cognition and biorobotics</b>	420
	The challenge of building a situated agent	421
	Situated cognition and knowledge representation	423
	Biorobotics: Insects and morphological computation	424
	<b>13.4 From subsumption architectures to behavior-based robotics</b>	430
	Subsumption architectures: The example of Allen	431
	Behavior-based robotics: TOTO	435
	Multi-agent programming: The Nerd Herd	438
<b>14</b>	<b>The cognitive science of consciousness</b>	445
	<b>14.1 The challenge of consciousness: Leibniz's Mill</b>	447
	<b>14.2 Consciousness and information processing: The Knowledge Argument</b>	448

<b>14.3 Information processing without conscious awareness: Some basic data</b>	<b>449</b>
Consciousness and priming	450
Non-conscious processing in blindsight and unilateral spatial neglect	453
<b>14.4 So what is consciousness for?</b>	<b>457</b>
What is missing in blindsight and spatial neglect	457
Milner and Goodale: Vision for action and vision for perception	458
What is missing in masked priming	463
<b>14.5 Two types of consciousness and the hard problem</b>	<b>463</b>
<b>14.6 The global workspace theory of consciousness</b>	<b>469</b>
The building blocks of global workspace theory	470
The global neuronal workspace theory	472
<b>14.7 Conclusion</b>	<b>475</b>
<b>15 Looking ahead: Challenges and applications</b>	<b>481</b>
<b>15.1 Exploring the connectivity of the brain: The connectome and the BRAIN initiative</b>	<b>482</b>
<b>15.2 Understanding what the brain is doing when it appears not to be doing anything</b>	<b>483</b>
<b>15.3 Building artificial brain systems?</b>	<b>484</b>
<b>15.4 Enhancing education</b>	<b>484</b>
<b>15.5 Building bridges to economics and the law</b>	<b>485</b>
<i>Glossary</i>	486
<i>Bibliography</i>	495
<i>Index</i>	514



BOXES

- 2.1 A conversation with ELIZA 32
- 3.1 What does each lobe do? 65
- 3.2 Brain vocabulary 66
- 4.1 The prisoner’s dilemma 104
- 6.1 Defining well-formed formulas in propositional logic 145
- 7.1 Calculating entropy 180
- 7.2 Calculating information gain 181
- 7.3 Calculating baseline entropy 183
- 7.4 Calculating the information gain for *Outlook?* 184
- 13.1 Basins of attraction in state space 410
- 14.1 A typical priming experiment 452



## FIGURES

- 1.1 A rat in a Skinner box  
Adapted from Spivey (2007). By permission of Oxford University Press, Inc 8
- 1.2 A 14-unit T-Alley maze  
Adapted from Elliott (1928) 9
- 1.3 A cross-maze, as used in Tolman, Ritchie, and Kalish (1946) 11
- 1.4 Schematic representation of a Turing machine  
Adapted from Cutland (1980) 15
- 1.5 A sample phrase structure tree for the sentence “John has hit the ball” 18
- 1.6 Donald Broadbent’s 1958 model of selective attention  
Adapted by courtesy of Derek Smith 20
- 2.1 A question for SHRDLU about its virtual micro-world  
Adapted from Winograd (1972) 33
- 2.2 An algorithm for determining whether a given input is a sentence or not  
Adapted from Winograd (1972) 35
- 2.3 Algorithms for identifying noun phrases and verb phrases  
Adapted from Winograd (1973) 36
- 2.4 Procedure for applying the concept CLEARTOP  
Adapted from Winograd (1972) 37
- 2.5 SHRDLU acting on the initial command to pick up a big red block  
Adapted from Winograd (1972: 8) 38
- 2.6 Instruction 3 in the SHRDLU dialog: “Find a block which is taller than the one you are holding and put it in the box.”  
Adapted from Winograd (1972: fig. 3) 39
- 2.7 Examples of the three-dimensional figures used in Shepard and Metzler’s 1971 studies of mental rotation  
Adapted from Shepard and Metzler (1971) 41
- 2.8 Depicting looking times as a function of angular rotation  
Adapted from Shepard and Metzler (1971) 42

- 2.9 Examples of vertically and horizontally oriented objects that subjects were asked to visualize in Kosslyn's 1973 scanning study  
Adapted from Kosslyn, Thompson, and Ganis (2006) 45
- 2.10 A table illustrating the three different levels that Marr identified for explaining information-processing systems  
From Marr (1982) 48
- 2.11 Example of conventional and unconventional perspectives  
Adapted from Warrington and Taylor (1973) 49
- 2.12 Two examples of Marr's primal sketch, the first computational stage in his analysis of the early visual system  
Adapted from Marr (1982) 51
- 2.13 An example of part of the 2.5D sketch  
Adapted from Marr (1982) 51
- 2.14 An illustration of Marr's 3D sketch, showing how the individual components are constructed  
Adapted from Marr (1982) 52
- 2.15 The place of the implementational level within Marr's overall theory  
Adapted from Marr (1982) 54
- 2.16 An illustration of which parts of the visual system are likely to be responsible for producing each of Marr's three stages  
From Prinz (2012) 55
- 3.1 The large-scale anatomy of the brain, showing the forebrain, the midbrain, and the hindbrain  
Adapted by courtesy of The Huntington's Disease Outreach Project for Education, at Stanford University 63
- 3.2 A vertical slice of the human brain, showing the cerebrum  
© TISSUEPIX/SCIENCE PHOTO LIBRARY 64
- 3.3 The division of the left cerebral hemisphere into lobes 64
- 3.4 The primary visual pathway 65
- 3.5 Image showing ventral stream and dorsal stream in the human brain visual system 67
- 3.6 Design and results of Ungerleider and Mishkin's crossed-lesion disconnection studies  
Adapted from Ungerleider and Mishkin (1982) 69
- 3.7 A generic three-layer connectionist network  
Adapted from McLeod, Plunkett, and Rolls (1998) 73



- 3.8** Gorman and Sejnowski's mine/rock detector network  
 Adapted from Gorman and Sejnowski (1988), printed in Churchland, Paul M., *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*, figure 10.2, page 203 ©1990, Massachusetts Institute of Technology, by permission of The MIT Press 75
- 3.9** Images showing the different areas of activation (as measured by blood flow) during the four different stages in Petersen *et al.*'s lexical access studies  
 From Posner and Raichle (1994) 80
- 3.10** A flowchart relating different areas of activation in Petersen *et al.*'s study to different levels of lexical processing  
 Reprinted by permission from Macmillan Publishers Ltd: Petersen *et al.* "Positron emission tomographic studies of the cortical anatomy of single-word processing," *Nature* (331), ©1988 81
- 4.1** Connections among the cognitive sciences, as depicted in the Sloan Foundation's 1978 report  
 Adapted from Gardner (1985) 89
- 4.2** Some of the principal branches of scientific psychology 92
- 4.3** Levels of organization and levels of explanation in the nervous system  
 Adapted from Shepherd (1994) 94
- 4.4** The spatial and temporal resolution of different tools and techniques in neuroscience  
 From Baars and Gage (2010) 96
- 4.5** The integration challenge and the "space" of contemporary cognitive science  
 Adapted by courtesy of David Kaplan 98
- 4.6** A version of the Wason selection task 101
- 4.7** Griggs and Cox's deontic version of the selection task 102
- 4.8** A microelectrode making an extracellular recording  
 Reproduced by courtesy of Dwight A. Burkhardt, University of Minnesota 107
- 4.9** Simultaneous microelectrode and fMRI recordings from a cortical site showing the neural response to a pulse stimulus of 24 seconds  
 Adapted from Bandettini and Ungerleider (2001) 109
- 5.1** Two illustrations of the neural damage suffered by the amnesic patient H.M.  
 Figure 1, "What's new with the amnesiac patient H.M.?" *Nature Neuroscience* 2002 Feb., 3(2): 153-60. 119
- 5.2** Baddeley's model of working memory 120
- 5.3** The initial stages of a functional decomposition of memory 121

- 5.4 A mechanism for detecting oriented zero-crossing segments  
Adapted from Marr and Hildreth (1980) 125
- 6.1 Allen Newell and Herbert Simon studying a search-space  
Reproduced by courtesy of Carnegie Mellon University Library 146
- 6.2 A typical traveling salesman problem 147
- 6.3 The structure of Fodor's argument for the language of thought hypothesis 159
- 6.4 Inside and outside the Chinese room  
Courtesy of Robert E. Horn, Visiting Scholar, Stanford University 161
- 7.1 A decision tree illustrating a mortgage expert system  
From Friedenber and Silverman (2006) 174
- 7.2 A graph illustrating the relation between entropy and probability in the context of  
drawing a ball from an urn 179
- 7.3 The first node on the decision tree for the tennis problem 185
- 7.4 The complete decision tree generated by the ID3 algorithm 185
- 7.5 A sample completed questionnaire used as input to an ID3-based expert system for  
diagnosing diseases in soybean crops  
Adapted from Michalski and Chilausky (1980) 187
- 7.6 Classifying different information-processing models of cognition 189
- 7.7 The basic architecture of WHISPER  
From Funt (1980) 190
- 7.8 The starting diagram for the chain reaction problem  
From Funt (1980) 192
- 7.9 The result of applying WHISPER's rotation algorithm in order to work out the trajectory  
of block B  
From Funt (1980) 193
- 7.10 The first solution snapshot output by WHISPER  
From Funt (1980) 194
- 7.11 The final snapshot representing WHISPER's solution to the chain reaction problem  
From Funt (1980) 194
- 7.12 A map of SHAKEY's physical environment  
From Nilsson (1984) 197
- 7.13 A labeled photograph of SHAKEY the robot  
Reproduced by courtesy of SRI International, Menlo Park, California 198
- 8.1 Schematic illustration of a typical neuron 213

- 8.2 An artificial neuron 214
- 8.3 Four different activation functions  
 Adapted from McLeod, Plunkett, and Rolls (1998) 215
- 8.4 Illustration of a mapping function 216
- 8.5 A single-layer network representing the Boolean function AND 219
- 8.6 A single-layer network representing the Boolean function NOT 220
- 8.7 The starting configuration for a single-layer network being trained to function as a NOT-gate through the perceptron convergence rule 222
- 8.8 Graphical representations of the AND and XOR (exclusive-OR) functions, showing the linear separability of AND 224
- 8.9 A multilayer network representing the XOR (exclusive-OR) function  
 Adapted from McLeod, Plunkett, and Rolls (1998) 226
- 8.10 The computational operation performed by a unit in a connectionist model  
 Adapted from McLeod, Plunkett, and Rolls (1998) 228
- 9.1 Pinker and Prince's dual route model of past tense learning in English 248
- 9.2 Rumelhart and McClelland's model of past tense acquisition  
 Adapted from Rumelhart, David E., James L. McClelland and PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition: Volume 1: Foundations*, figure 4, page 242, ©1986 Massachusetts Institute of Technology, by permission of The MIT Press 250
- 9.3 Performance data for Rumelhart and McClelland's model of past tense learning  
 Adapted from Rumelhart, David E., James L. McClelland and PDP Research Group, *Parallel Distributed Processing: Explorations in the Microstructures of Cognition: Volume 1: Foundations*, figure 1, page 22, ©1986 Massachusetts Institute of Technology, by permission of The MIT Press 251
- 9.4 The network developed by Plunkett and Marchman to model children's learning of the past tense  
 Adapted from Plunkett and Marchman (1993) 252
- 9.5 A comparison of the errors made by a child and by the Plunkett-Marchman neural network model of tense learning  
 Adapted from McLeod, Plunkett, and Rolls (1998) 253
- 9.6 Schematic representation of the habituation and test conditions in Baillargeon's drawbridge experiments  
 Baillargeon 1987. Copyright ©1987 by the American Psychological Association. Reproduced with permission 256

- 9.7** Schematic representation of an experiment used to test infants' understanding of Spelke's principle of cohesion  
 Adapted from Spelke and Van de Walle (1993) 257
- 9.8** Schematic representation of an experiment testing infants' understanding of the principle that only surfaces in contact can move together  
 Adapted from Spelke and Van de Walle (1993) 258
- 9.9** Schematic depiction of events that accord with, or violate, the continuity or solidity constraints  
 Adapted from Spelke and Van de Walle (1993) 259
- 9.10** A series of inputs to the network as a barrier moves in front of a ball and then back to its original location  
 Adapted from Munakata, Y., McClelland, J. L., Johnson, M. H., Seigler, R. S. (1997)  
 Copyright ©1997 by the American Psychological Association.  
 Adapted with permission 263
- 9.11** Recurrent network for learning to anticipate the future position of objects  
 Adapted from Munakata *et al.* (1997) 265
- 9.12** A balance beam 266
- 9.13** The architecture of the McClelland and Jenkins network for the balance beam problem  
 Elman, Jeffrey, Elizabeth Bates, Mark H. Johnson, Annette Karmiloff-Smith, Domenico Parisi, and Kim Plunkett, *Rethinking Innateness: A Connectionist Perspective on Development*, figure 3.19, © 1996 Massachusetts Institute of Technology, by permission of The MIT Press 267
- 10.1** The architecture of a simple reflex agent  
 Adapted from Russell and Norvig (2009) 282
- 10.2** The architecture of a goal-based agent  
 Adapted from Russell and Norvig (2009) 283
- 10.3** The architecture of a learning agent  
 Russell, Stuart; Norvig, Peter, *Artificial Intelligence: A Modern Approach*, 3rd Edition, © 2010, pp. 49, 52, 55. Adapted by permission of Pearson Education, inc., Upper Saddle River, NJ 284
- 10.4a** Franz Joseph Gall (1758-1828)  
 Courtesy of the Smithsonian Institution Libraries, Washington DC 286
- 10.4b** A three-dimensional model of Gall's phrenological map developed by the American phrenologist Lorenzo Niles Fowler (1811-96)  
 Reproduced courtesy of the Science Museum/Science & Society Picture Library 287
- 10.4c** Jerry Fodor (1935-) 288

10.5	The evolutionary biologist W. D. Hamilton (1936–2000). © Jeffrey Joy	299
10.6	The ACT-R/PM cognitive architecture Courtesy of Lorin Hochstein, University of Southern California	307
11.1	Luria’s 1970 diagram of the functional organization of the brain Adapted from Luria (1970)	317
11.2	Map of the anatomy of the brain showing the four lobes and the Brodmann areas Reproduced courtesy of Applied Neuroscience Inc.	320
11.3	A connectivity matrix for the visual system of the macaque monkey Adapted from Felleman and Van Essen (1991)	322
11.4	An anatomical wiring diagram of the visual system of the macaque monkey Adapted from Felleman and Van Essen (1991)	323
11.5	The results of single-neuron recordings of a mirror neuron in area F5 of the macaque inferior frontal cortex Adapted from Iacoboni and Dapretto (2006)	326
11.6	Typical patterns of EEG waves, together with where/when they are typically found Reprinted from Baars, Bernard J., and Gage, Nicole M., <i>Fundamentals of Cognitive Neuroscience: A Beginner’s Guide</i> , © 2012, with permission from Elsevier.	328
11.7a	Common experimental design for neurophysiological studies of attention	333
11.7b	Example of occipital ERPs recorded in a paradigm of this nature	333
11.7c	Single-unit responses from area V4 in a similar paradigm	334
11.7d	Single-unit responses from area V1 showing no effect of attention Adapted from Luck and Ford (1998), with permission from <i>Neuroimaging for Human Brain Function</i> Copyright ©1998 National Academy of Sciences, courtesy of the National Academies Press, Washington DC	334
11.8	Frontoparietal cortical network during peripheral visual attention Gazzaniga, Michael, ed., <i>The New Cognitive Neurosciences</i> , second edition, Plates 30 & 31, © 1999 Massachusetts Institute of Technology, by permission of The MIT Press	339
11.9	Schematic of delayed saccade task Adapted from R. L. White and L. H. Snyder, “Subthreshold microstimulation in frontal eye fields updates spatial memories,” <i>Experimental Brain Research</i> 181, 477–92 © Springer-Verlag 2007	340
11.10	Peripheral attention vs. spatial working memory vs. saccadic eye movement across studies Gazzaniga, Michael, ed., <i>The New Cognitive Neurosciences</i> , second edition, Plates 30 & 31, © 1999 Massachusetts Institute of Technology, by permission of The MIT Press	342

- 12.1 An example of metarepresentation 357
- 12.2 The general outlines of Leslie's model of pretend play  
Adapted from Leslie (1987) 358
- 12.3 Leslie's Decoupler model of pretense  
Adapted from Leslie (1987) 359
- 12.4 The task used by Baron-Cohen, Leslie, and Frith to test for children's understanding of false belief  
Adapted from Baron-Cohen, Leslie, and Frith (1985) 363
- 12.5 The connection between pretend play and success on the false belief task  
Adapted from the Open University OpenLearn Unit DSE232\_1, courtesy of the Open University 365
- 12.6 Baron-Cohen's model of the mindreading system 370
- 12.7 What goes on in representing belief 379
- 12.8 What goes on in representing perception 379
- 12.9 A schematic version of standard simulationism  
Adapted from Nichols *et al.* (1996) 383
- 12.10 Schematic representation of brain regions associated with the attribution of mental states  
Adapted from Saxe, Carey, and Kanwisher (2004) 388
- 12.11 Schematic overview of the frontoparietal mirror neuron system (MNS) and its main visual input in the human brain  
Adapted from Iacoboni and Dapretto (2006) 393
- 13.1 The trajectory through state space of an idealized swinging pendulum  
By permission of M. Casco Associates 406
- 13.2 The state space of a swinging pendulum in a three-dimensional phase space  
By permission of M. Casco Associates 407
- 13.3 Illustration of the Watt governor, together with a schematic representation of how it works  
Adapted from Bechtel 1998, *A Companion to Cognitive Science* 409
- 13.4 An example of a computational model of motor control  
Adapted from Shadmehr and Krakauer (2008) 413
- 13.5 The stage IV search task, which typically gives rise to the A-not-B-error in infants at around 9 months  
Adapted from Bremner (1994) 415

- 13.6 An infant sitting for an A trial (left) and standing for a B trial (right)  
 Adapted from Smith and Thelen (2003) 417
- 13.7 Applying the dynamical field model to the A-not-B error  
 Figure 2 in Smith and Thelen, *Development as a Dynamic System*, Elsevier 2003 418
- 13.8 The organizing principles of biorobotics  
 Reproduced courtesy of Dimitrios Lambrinos 425
- 13.9 The anatomy of a cricket, showing the different routes that a sound can take to each ear  
 Adapted from Clark (2001) 426
- 13.10 A robot fish called WANDA  
 Reproduced courtesy of Marc Zeigler, University of Zurich 428
- 13.11 WANDA swimming upwards  
 From Pfeifer, Iida, and Gómez (2006) 429
- 13.12 Yokoi's robot hand  
 Reproduced courtesy of Gabriel Gómez, Alejandro Hernandez Arieta, Hiroshi Yokoi, and Peter Eggenberger Hotz, University of Zurich 430
- 13.13 The Yokoi hand grasping two very different objects  
 From Pfeifer, Iida, and Gómez (2006) 431
- 13.14 Rodney Brooks's robot Allen, his first subsumption architecture robot  
 Adapted from Haugeland, John, ed., *Mind Design II: Philosophy, Psychology, and Artificial Intelligence*, figures 15.1 & 15.2, © 1997 Massachusetts Institute of Technology, by permission of The MIT Press 432
- 13.15 The layers of Allen's subsumption architecture  
 Adapted from Haugeland, John, ed., *Mind Design II: Philosophy, Psychology, and Artificial Intelligence*, figures 15.1 & 15.2, © 1997 Massachusetts Institute of Technology, by permission of The MIT Press 433
- 13.16 The Nerd Herd, together with the pucks that they can pick up with their grippers  
 Adapted from Haugeland, John, ed., *Mind Design II: Philosophy, Psychology, and Artificial Intelligence*, figures 15.1 & 15.2, © 1997 Massachusetts Institute of Technology, by permission of The MIT Press 439
- 14.1 A typical congruence priming experiment  
 From Finkbeiner and Forster (2008) 451
- 14.2 Deficits found in patients with left spatial neglect  
 From Driver and Vuilleumier (2001) 454
- 14.3 Form perception in blindsight  
 From Trevethan, Sahraie, and Weiskrantz (2007) 456

- 14.4 Non-conscious perception in neglect  
From Marshall and Halligan (1988) 457
- 14.5 Visuomotor and perceptual matching tasks  
From Milner and Goodale (1998) 460
- 14.6 Grasping and the Ebbinghaus illusion  
Reprinted from *Current Biology*, S. Salvatore Aglioti, Joseph F. X. DeSouza, Melvyn and A. Goodale, 'Size-contrast illusions deceive the eye but not the hand' 679–685, © 1995, with permission from Elsevier. 461
- 14.7 Fang and He's interocular suppression task  
Adapted from Fang and He (2005) 462
- 14.8 Three versions of the global workspace theory  
From Dehaene and Changeux (2001) 473
- 14.9 The neural substrates of the global workspace  
From Dehaene and Naccache (2001), Figure 3 on p. 27 474





TABLES

- 7.1 SHAKEY’S five levels 199
- 7.2 How SHAKEY represents its own state 200
- 7.3 SHAKEY’s intermediate-level routines 202
- 9.1 The stages of past tense learning according to verb type 247
- 10.1 Why we cannot use the language of thought hypothesis to understand central processing: A summary of Fodor’s worries 294
- 10.2 Comparing the symbolic and subsymbolic dimensions of knowledge representation in the hybrid ACT-R/PM architecture 310  
From Lovett and Anderson (2005)
- 11.1 Comparing techniques for studying connectivity in the brain 330
- 12.1 The three groups studied in Baron-Cohen, Leslie, and Frith 1985 362
- 13.1 The five basis behaviors programmed into Mataric’s Nerd Herd robots 439



## PREFACE



### About this book

There are few things more fascinating to study 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, there are very few academic areas that 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.

This enormous range of information out there about the mind can be overwhelming, both for students and for instructors. I had direct experience of how challenging this can be when I was Director of the Philosophy-Neuroscience-Psychology program at Washington University in St. Louis. My challenge was to give students a broad enough base while at the same time bringing home that cognitive science is a field in its own right, separate and distinct from the disciplines on which it draws. I set out to write this book because my colleagues and I were unable to find a book that really succeeds in doing this.

Different textbooks have approached this challenge in different ways. Some 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 learnt much from these books and they can be very helpful. But I often have the feeling that students 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, in order to give students a sense of cognitive science's breadth and range.

Cognitive science has only been with us for forty or so years. But in that time it has changed a lot. At one time cognitive science was associated with the idea that we can understand the mind without worrying about its biological machinery – we can understand the software without understanding the hardware, to use a popular image. But this is now really a minority view. Neuroscience is now an absolutely fundamental part of cognitive science. Unfortunately this has not really been reflected in textbooks on cognitive science. This book presents a more accurate picture of how central neuroscience is to cognitive science.



## How the book is organized

This book is organized into five parts.



### Part I: Historical overview

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: The integration challenge

The two chapters in Part II bring out what is distinctive about cognitive science. They do this in terms of what I call the integration challenge. This is the challenge of developing a unified framework that makes explicit the relations between the different disciplines on which cognitive science draws and the different levels of organization that it studies. In Chapter 4 we look at two examples of *local integration*. The first example explores how evolutionary psychology has been used to explain puzzling data from human decision-making, while the second focuses on what exactly it is that is being studied by techniques of neuro-imaging such as functional magnetic resonance imaging (fMRI).

In Chapter 5 I propose that one way of answering the integration challenge is through developing models of mental architecture. A model of mental architecture includes

- 1 an account of how the mind is organized into different cognitive systems, and
- 2 an account of how information is processed in individual cognitive systems.

This approach to mental architecture sets the agenda for the rest of the book.



## Part III: Information-processing models of the mind

The four chapters in Part III explore the two dominant models of information processing in contemporary cognitive science. The first model 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. The theoretical case for the physical symbol system hypothesis is discussed in Chapter 6, while Chapter 7 gives three very different examples of research within that paradigm – from data mining, artificial vision, and robotics.

The second model of information processing derives from models of artificial neurons in computational neuroscience and connectionist artificial intelligence. Chapter 8 explores the motivation for this approach and introduces some of the key concepts, while Chapter 9 shows how it can be used to model aspects of language learning and object perception.



## Part IV: How is the mind organized?

A mental architecture includes a model both of information processing and of how the mind is organized. The three chapters in Part IV look at different ways of tackling this second problem. Chapter 10 examines the idea that some forms of information processing are carried out by dedicated cognitive modules. It looks also at the radical claim, proposed by evolutionary psychologists, that the mind is simply a collection of specialized modules. In Chapter 11 we look at how techniques such as functional neuroimaging can be used to study the organization of the mind. Chapter 12 shows how the theoretical and methodological issues come together by working 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).



## Part V: New horizons

As emerges very clearly in the first four parts of the book, cognitive science is built around some very basic theoretical assumptions – and in particular around the assumption that the mind is an information-processing system. In Chapter 13 we look at two ways in which cognitive scientists have proposed extending and moving beyond this basic assumption. One of these research programs is associated with the dynamical systems hypothesis in cognitive science. The second is opened up by the situated/embodied cognition movement. Chapter 14 explores 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.



Using this book in courses

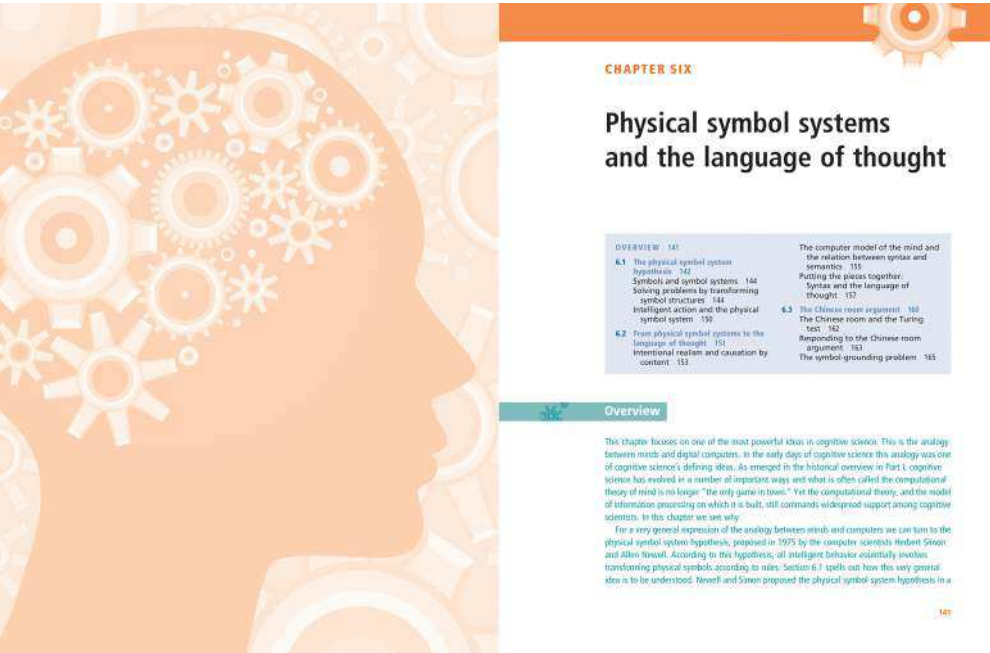
This book has been designed to serve as a self-contained text for a single semester (12-15 weeks) introductory course on cognitive science. Students taking this course may have taken introductory courses in psychology and/or philosophy, but no particular prerequisites are assumed. All the necessary background 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:

- **Part-openers and chapter overviews** The book is divided into five parts, as described above. Each part begins with a short introduction to give the reader a broad picture of what lies ahead. Each chapter begins with an overview to orient the reader.



- **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).

216

Information processing in neural networks

6.2

Single-layer networks and Boolean functions

One way of thinking about information processing is in terms of mapping functions. Functions are being understood here in the strict mathematical sense. The basic idea of a function should be familiar, even if the terminology may not be. Addition is a function. Given two numbers as inputs, the addition function yields a third number as output. The output is the sum of the two inputs. Multiplication is also a function. Here the third number is the product of the two inputs.

Let us make this a little more precise. Suppose that we have a set of items. We can call that a domain. Let there be another set of items, which we can call the range. A mapping function maps each item from the domain onto exactly one item from the range. The defining feature of a function is that no item in the domain gets mapped to more than one item in the range. Functions are single-valued. The operation of taking square roots, for example, is not a function (at least when negative numbers are included, since every positive number has two square roots).

**Exercise 6.1** Give another example of an arithmetical operation that counts as a function. And another example of an operation that is not a function.

Figure 6.4 gives an example of a mapping function. The arrows indicate which item in the domain is mapped to each item in the range. It is perfectly acceptable for two or more items in the domain to be mapped to a single item in the range (as is the case with  $A_1$  and  $A_2$ ). But, because functions are single-valued, no item in the domain can be mapped onto more than one item in the range.

The mapping function of addition has a domain made up of all the possible pairs of numbers. Its range is made up of all the numbers. In this case we can certainly have several different items in the domain mapping onto a single item in the range. Take  $A_3$  to be the pair (1, 3) and  $A_2$  to be the pair (2, 2). The addition function maps both  $A_3$  and  $A_2$  onto 4 (which we can take to be  $B_4$ ).

Consider now a mapping function with two items in its range. We can think about this as a way of classifying objects in the domain of the function. Imagine that the

$A_1$

$A_2$

$A_3$

$A_4$

$B_1$

$B_2$

$B_3$

$B_4$

Domain Range

**Figure 6.4** Illustration of a mapping function. A mapping function maps each item in its domain to exactly one item in its range.

217

Single-layer networks and Boolean functions

domain of the function contains all the natural numbers and the range of the function contains two items corresponding to TRUE and FALSE. Then we can identify any subset we please of the natural numbers by mapping the members of that subset onto TRUE and all the others onto FALSE. If the subset that the function maps onto TRUE contains all and only the even numbers, for example, then we have a way of picking out the set of the even numbers. This in fact is how the famous mathematician Gottlob Frege, who invented modern logic, thought about concepts. He thought of the concept even number as a function that maps every even number to TRUE, and everything else to FALSE.

Anybody who has taken a course in elementary logic will be familiar with an important class of mapping functions. These functions all have the same range as our even number function – namely, the set consisting of the two truth values TRUE and FALSE. Using the standard notation for sets we can write the range of the function as {TRUE, FALSE}. Instead of having numbers in the domain, however, the domain of these functions is made up of pairs of truth values. These functions, the so-called *binary Boolean functions*, take pairs of truth values as their inputs and deliver truth values as their outputs. They are called binary functions because the domain of the function consists of pairs (addition is also a binary function). They are called Boolean functions (after the nineteenth-century mathematician George Boole) because both the domain and range are built up from truth values.

**Exercise 6.2** Give an example of a unary (one-place) Boolean function and an example of a ternary (three-place) Boolean function.

There are four different possible pairs of truth values. These pairs form the domain of the binary Boolean functions. The range, as with all Boolean functions, is given by the set {TRUE, FALSE}, as illustrated below:

DOMAIN	RANGE
FALSE, FALSE	FALSE
FALSE, TRUE	FALSE
TRUE, FALSE	TRUE
TRUE, TRUE	TRUE

Each binary Boolean function assigns either TRUE or FALSE to each pair of truth values.

It is easier to see what is going on if you think of a binary Boolean function as a way of showing how the truth value of a complex sentence is determined by the truth values of the individual sentences from which they are built up. Some of the Boolean functions should be very familiar. There is a binary Boolean function standardly known as AND, for example. AND maps the pair {TRUE, TRUE} to TRUE and maps all other pairs of truth

■ **Boxes and optional material** Boxes have been included to provide further information about the theories and research discussed in the text. Some of the more technical material has been placed in boxes that are marked optional. Readers are encouraged to work through these, but the material is not essential to the flow of the text.

218

Applying the symbolic paradigm

219

An algorithm for recursive learning

BOX 7.1

Calculating entropy

OPTIONAL

BOX 7.2

Calculating information gain

OPTIONAL

Entropy in the information-theoretic sense is a way of measuring uncertainty. How do we turn this intuitive idea into a mathematical formula?

To keep things simple we will just calculate the entropy of a set of examples relative to a binary attribute. A binary attribute is one that has two possible values. The example in the text of blood? is a binary attribute, for example. We need some notation – as follows:

$S$  the set of examples.  
 $|S|$  the number of examples in  $S$ .  
 $A$  the (binary) attribute.  
 $N(A^{(+)})$  the number of examples with attribute  $A$ .  
 $N(A^{(-)})$  the number of examples lacking attribute  $A$ .

So, the proportion of examples in  $S$  with attribute  $A$  is given by  $\frac{N(A^{(+)})}{|S|}$ , and the proportion of examples in  $S$  lacking attribute  $A$  is given by  $\frac{N(A^{(-)})}{|S|}$ . If we abbreviate these by  $\text{Prop}(A^{(+)})$  and  $\text{Prop}(A^{(-)})$  respectively, then we can calculate the entropy of  $S$  relative to  $A$  with the following equation:

$$\text{Entropy } S/A = -\text{Prop}(A^{(+)}) \log_2 \text{Prop}(A^{(+)}) - \text{Prop}(A^{(-)}) \log_2 \text{Prop}(A^{(-)})$$

This is not as bad as it looks! We are working in base 2 logarithms because we are dealing with a binary attribute.

**Exercise** To make sure that you are comfortable with this equation, refer to the example in the text and check:

(a) that the entropy is 1 when the proportion of black balls is 0.5.  
(b) that the entropy is 0.88 when the proportion of black balls is 0.7.

**NB** Your calculator may not be able to calculate logarithms to the base 2 directly. The log function will most likely be base 10. You may find the following formula helpful:  $\log_2(x) = \log_{10}(x) / \log_{10}(2)$  for any base.

The baseline entropy of  $S^*$  relative to the target attribute. This is the starting point from which it can then calculate which of the remaining attributes has the highest information gain. The attribute with the highest information gain is selected and assigned to the node.

This process is repeated until each branch of the tree ends in a value for the target attribute. This will happen if the attributes on a particular branch end up narrowing the set of examples down so that they all have the same value for the target attribute. When every branch is closed in this way the algorithm halts.

We can measure information gain once we have a way of measuring entropy. Assume that we are starting at a node on the tree. It may be the starting node, but need not be. The node has associated with it a particular set  $S^*$  of examples. If the node is the starting node then  $S^*$  will contain all the examples – i.e. we will have  $S^* = S$ . If the node is further down the tree then it will be some subset of  $S$  – i.e. we have  $S^* \subset S$ .

The first step is to calculate the entropy of  $S^*$  relative to the target attribute  $A$  – i.e.  $\text{Entropy}(S^*/A)$ . This can be done using the formula in Box 7.1 and gives the algorithm its baseline again.

Now what we want to do is to calculate how much that uncertainty would be reduced if we had information about whether or not the members of  $S^*$  have a particular attribute – say,  $B$ .

So, the second step is to calculate the entropy with respect to the target attribute of the subset of  $S^*$  that has attribute  $B$  – what according to the notation we used in Box 7.1 we call  $B^{(+)}$ . This can be done using the formula from Box 7.1 to give a value for  $\text{Entropy}(B^{(+)}/A)$ .

The third step is the same as the second, but in this case we calculate the entropy of  $B^{(-)}$  with respect to the target attribute – i.e. the subset of  $S^*$  that does not have attribute  $B$ . This gives a value for  $\text{Entropy}(B^{(-)}/A)$ .

Finally, the algorithm puts these together to work out the information gain in  $S^*$  due to attribute  $B$ . This is given by the following formula:

$$\text{Gain}(S^*, B) = \text{Entropy}(S^*/A) - \text{Prop}(B^{(+)}) \times \text{Entropy}(B^{(+)}/A) - \text{Prop}(B^{(-)}) \times \text{Entropy}(B^{(-)}/A)$$

As in Box 7.1,  $\text{Prop}(A^{(+)})$  stands for the proportion of  $S^*$  that has attribute  $A$ .

**ID3 in action**

We can illustrate how ID3 works by showing how it can produce a decision tree for solving a relatively simple problem – deciding whether or not the weather is suitable for playing tennis. In order to apply ID3 we need a database. So imagine that, as keen tennis players who seriously consider playing tennis every day, we collect information for two weeks. For each day we log the principal meteorological data and note whether or not we decide to play tennis on that day.

The target attribute is Play Tennis? The other attributes are the general weather outlook, the temperature, the humidity, and the wind. Here they are with the values they can take.