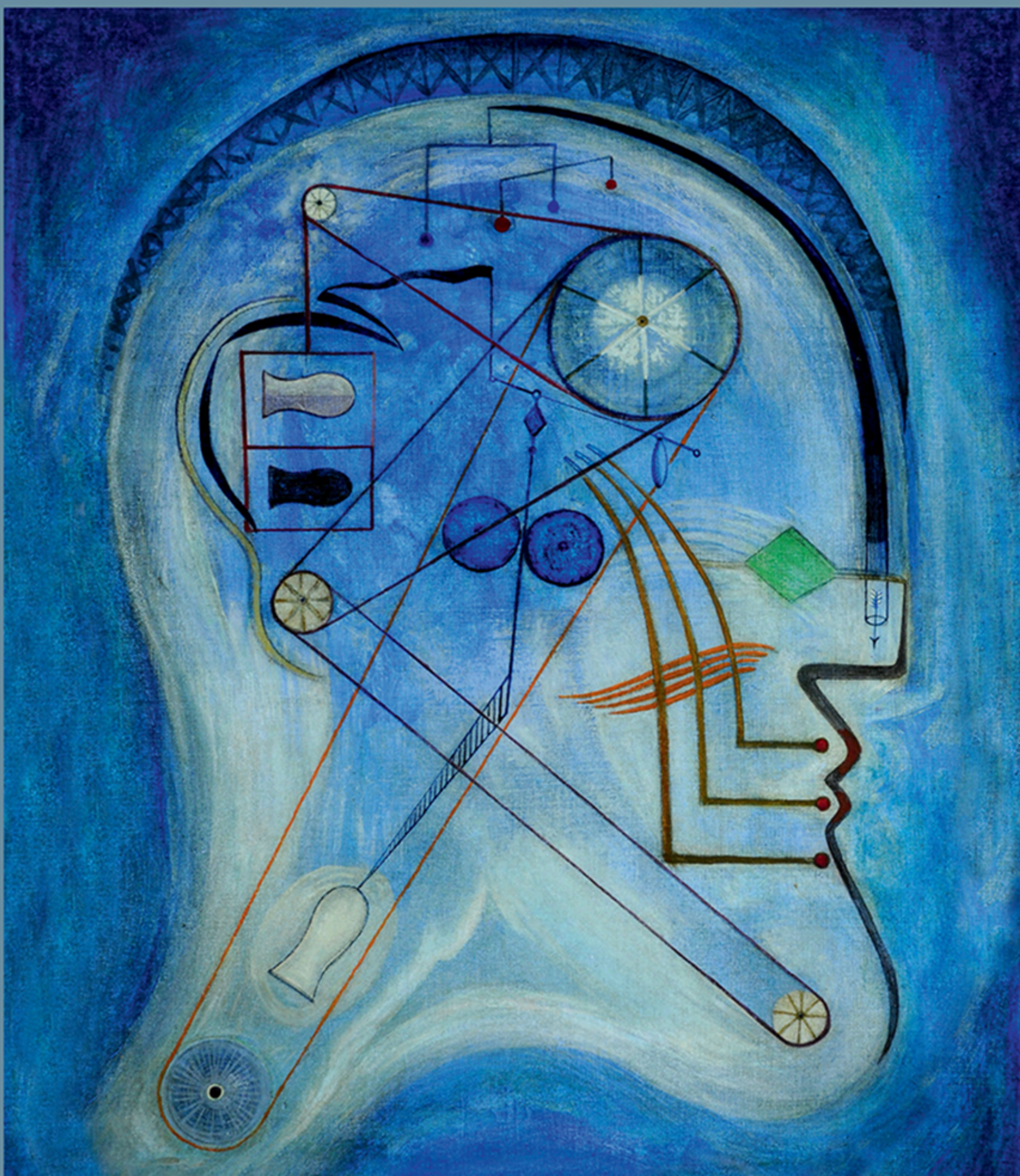
José Luis Bermúdez

Cognitive Science

An Introduction to the Science of the Mind

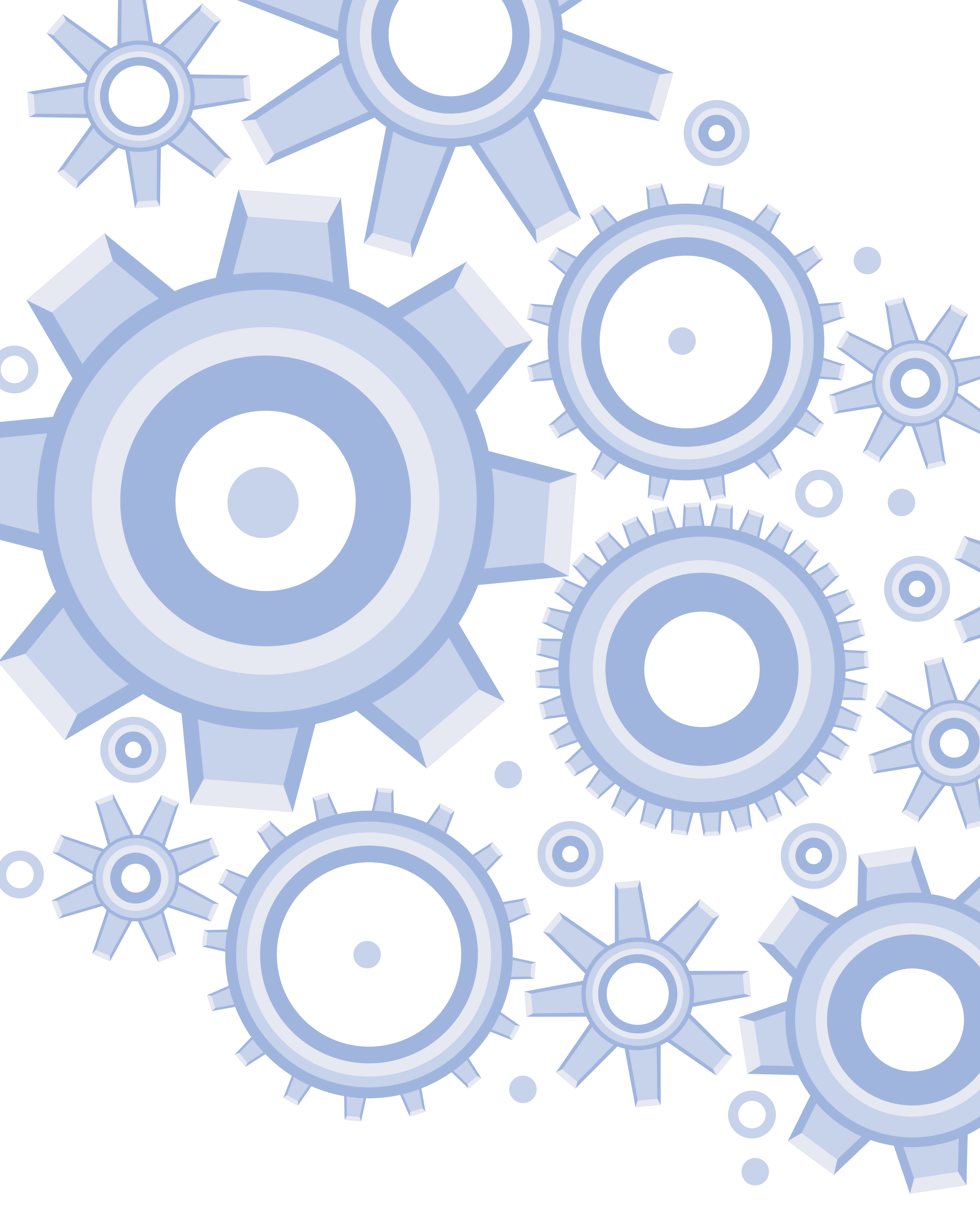
Third Edition



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

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

CONTENTS

CONTENTS

BOXES

FIGURES

TABLES

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.

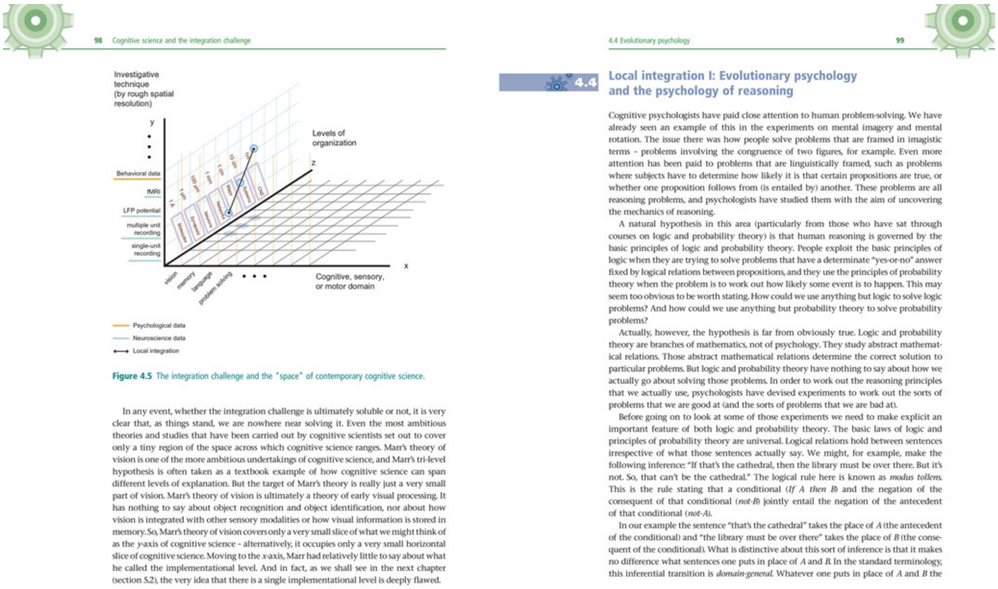
Using This Book in Courses

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.



98 Cognitive science and the integration challenge

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 tiny region of the space across which cognitive science ranges. 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 early visual processing. It has nothing to say about object recognition and object identification, nor about how vision is integrated with other sensory modalities or how visual information is stored in memory. So, Marr's theory of vision covers only a very small slice of what we might think of as the *y*-axis 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 99

4.4 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. The issue there was how people solve problems that are framed in imagistic terms – problems involving the congruence of two figures, for example. Even 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.

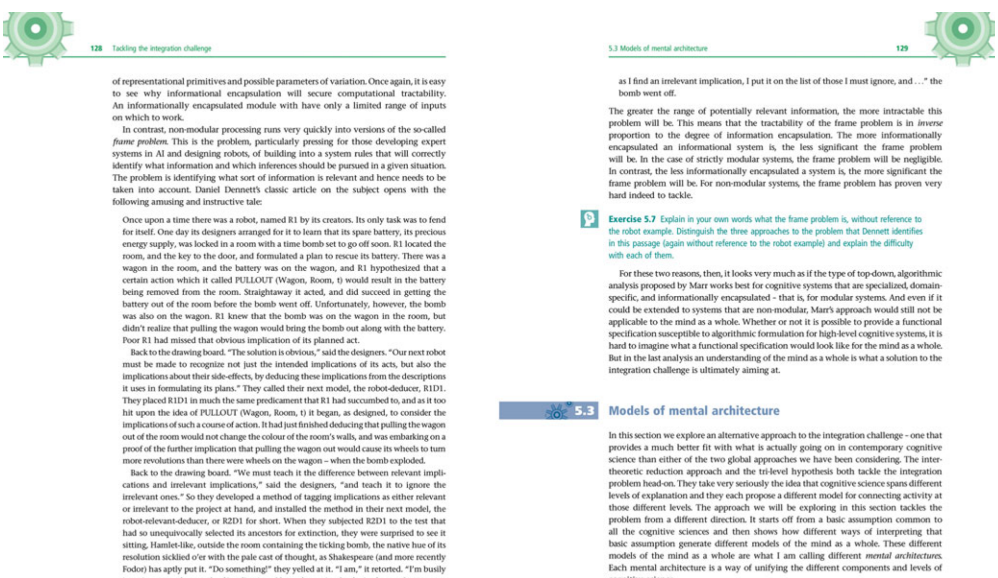
A natural hypothesis in this area (particularly from those who have sat through courses on logic and probability theory) is that human reasoning is governed by the basic principles of logic and probability theory. People exploit the basic principles of logic when they are trying to solve problems that have a determinate "yes-or-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 actually use, psychologists have devised experiments to work out the sorts of problems that we are good at (and the sorts of problems that we are bad at).

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, that can't be the cathedral." The logical rule here is known as *modus tollens*. This is the rule stating 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).



128 Tackling the integration challenge

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, non-modular processing runs very quickly into versions of the so-called *frame problem*. This is the problem, particularly pressing for those developing expert systems in 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 fend 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 the battery was on the wagon, and R1 hypothesized that a certain action which it called PULLOUT (Wagon, Room, t) 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 missed that obvious implication of its planned act.
>
> 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-deducer, 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, t) it began, as designed, to consider the implications of such a course of action. It had just finished deducing that pulling the wagon out of the room would not change the colour of the room's walls, and was embarking 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 method of tagging implications as either relevant or irrelevant to the project at hand, and installed the method in their next model, the robot-relevant-deducer, 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 sicklied 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 129

> 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 analysis proposed by Marr works best for cognitive systems that are specialized, domain-specific, and informationally encapsulated – that is, for modular systems. And even if it could 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 explanation and they each propose a different model for connecting activity at those different levels. 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.

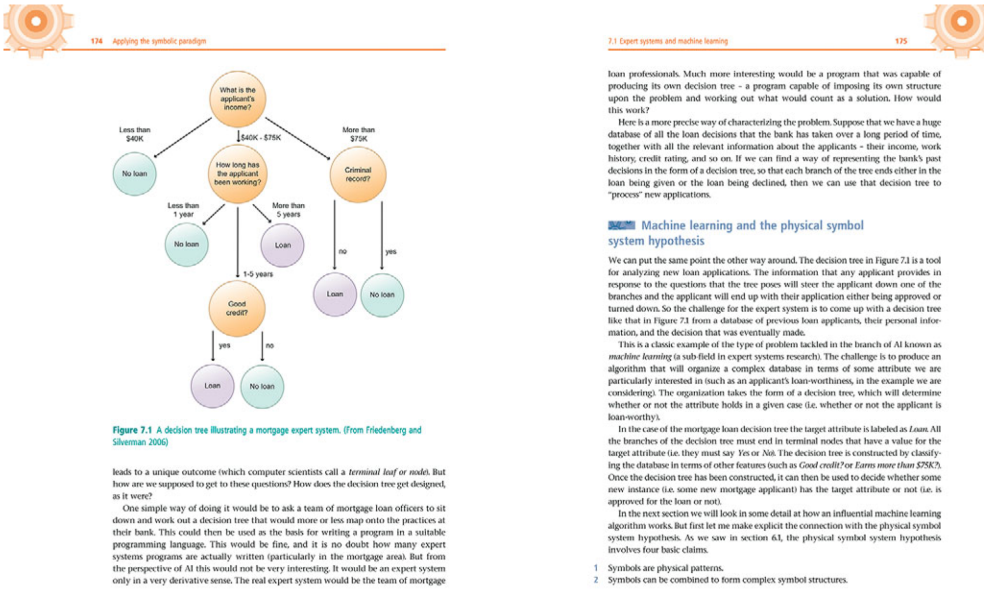


Figure 7.1 A decision tree illustrating a mortgage expert system. (From Friedenberg and Silverman 2006)

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 could then be used as the basis for writing a program in a suitable programming language. This would be fine, and it is no doubt how 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 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 work?

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 like that in Figure 7.1 from a database of previous loan applicants, their personal information, and the decision that 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 $75K?*). Once the decision tree has been constructed, it can then be used to decide whether some new instance (i.e. some new 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

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 III 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

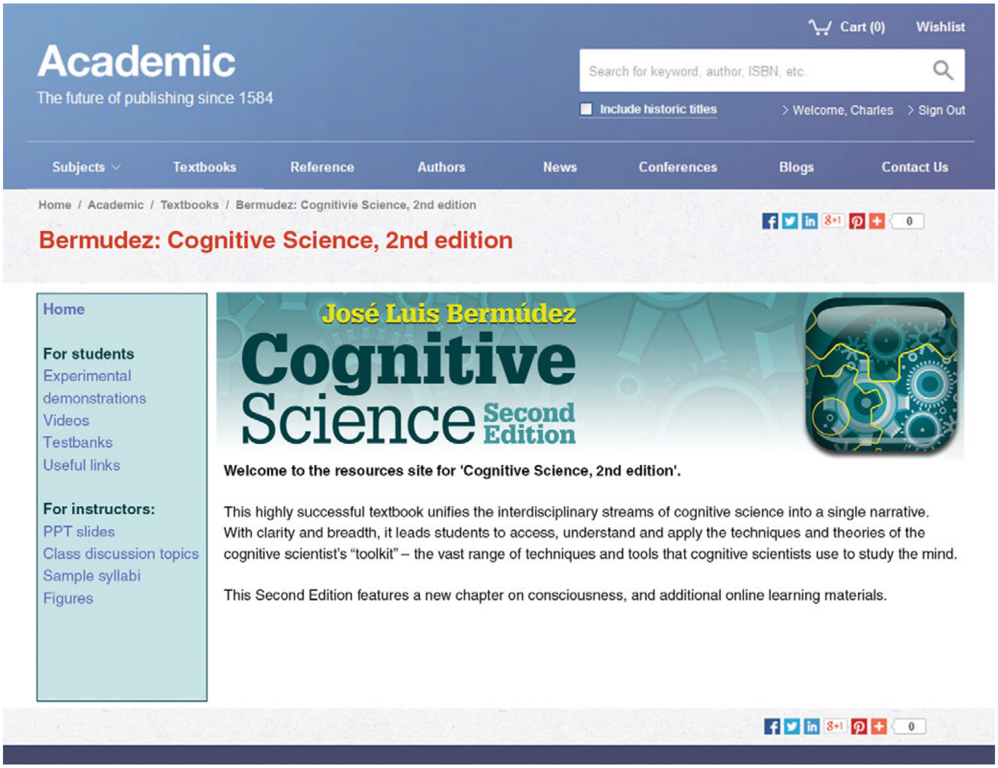
A course website accompanies the book. It can be found at www.cambridge.org/bermudez3. This website contains

- 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.



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

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

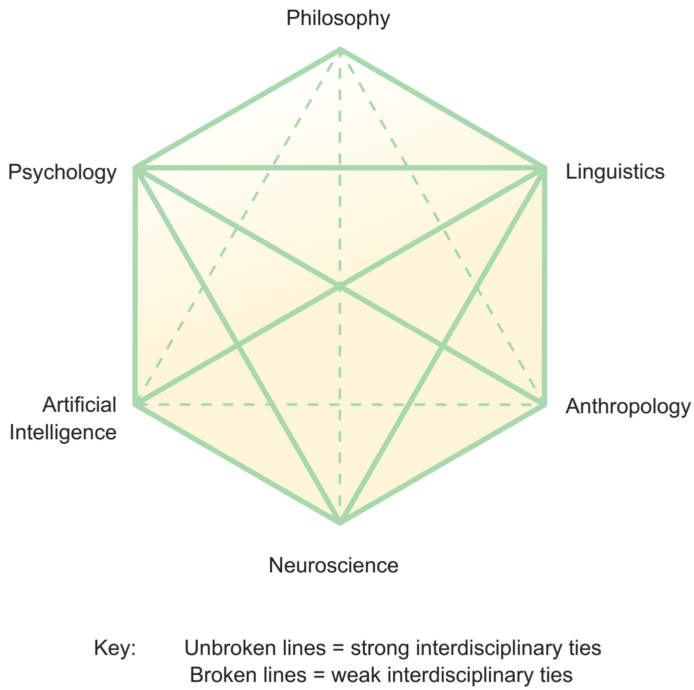


Figure 0.1 Connections among the cognitive sciences, as depicted in the Sloan Foundation's 1978 report. Unbroken lines indicate strong interdisciplinary links, while broken lines indicate weaker links. (Adapted from Gardner 1985)

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 subbranches all study different aspects of mind and behavior at the level of the organism.

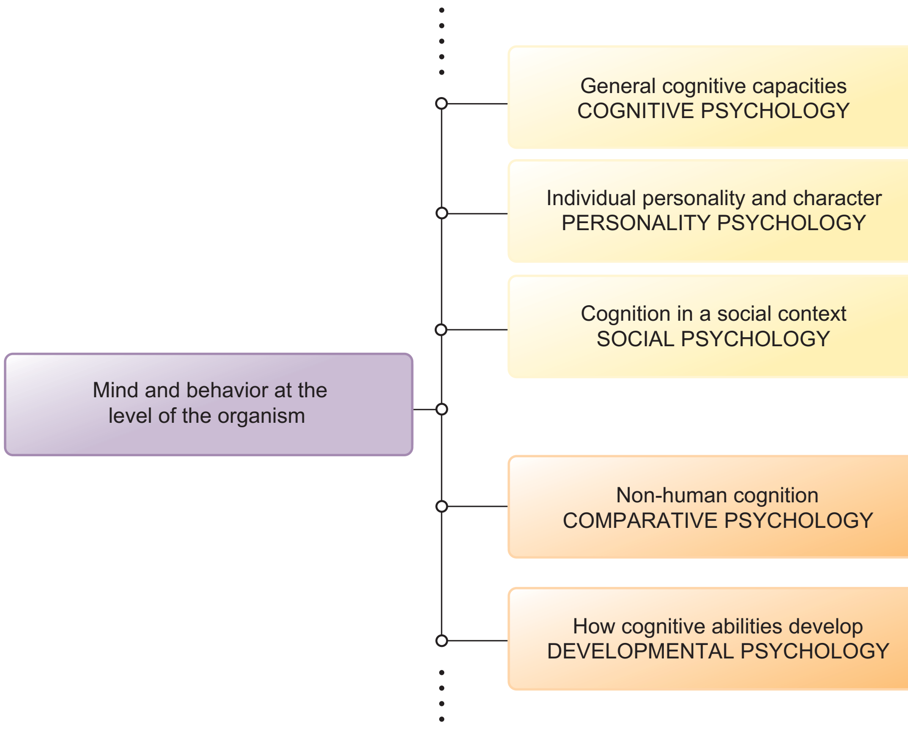


Figure 0.2 Some of the principal branches of scientific psychology.

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

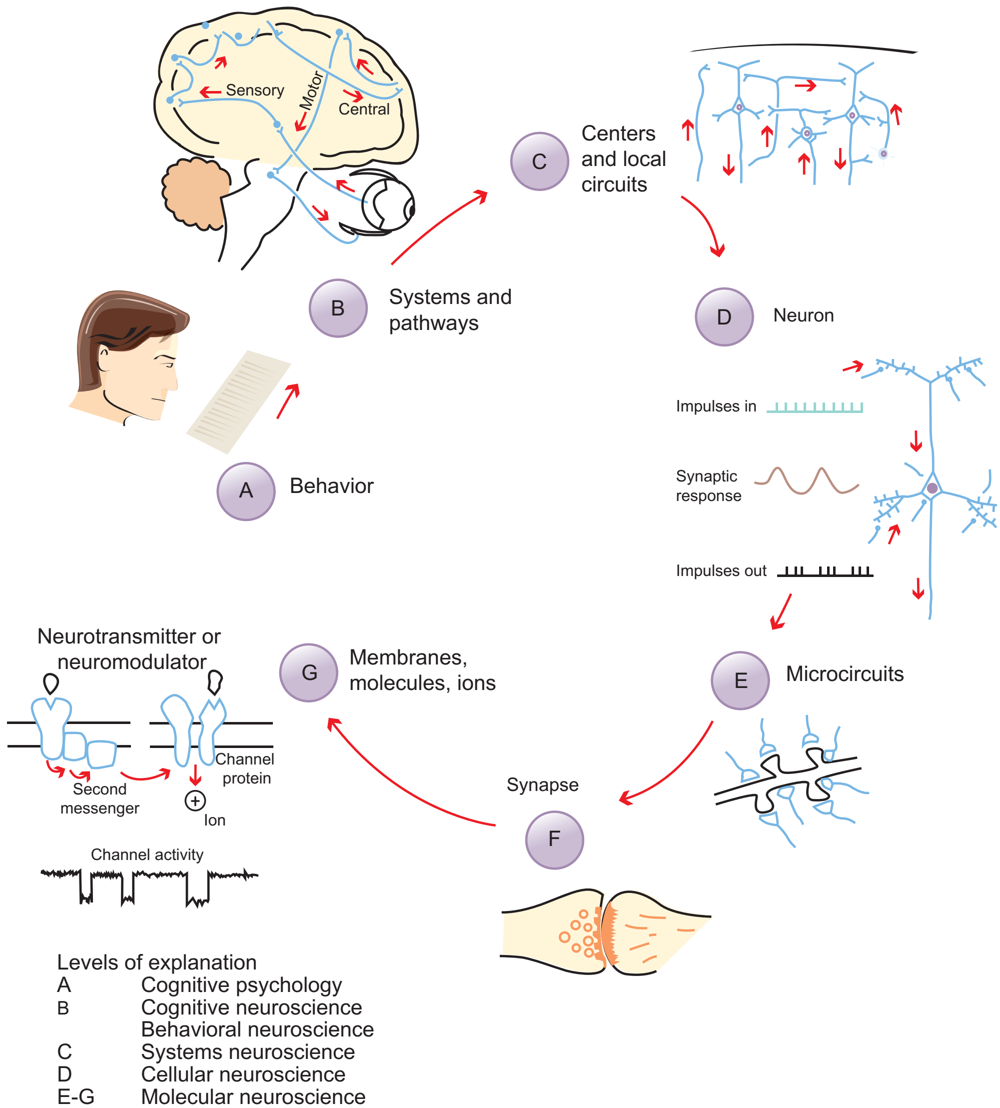


Figure 0.3 Levels of organization and levels of explanation in the nervous system. (Adapted from Shepherd 1994)

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).

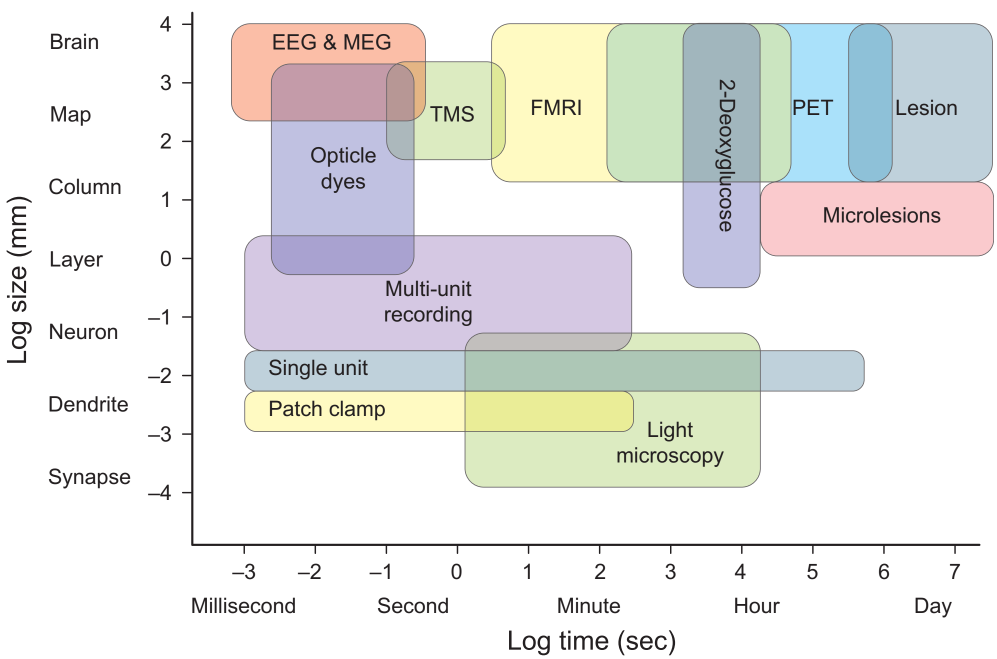


Figure 0.4 The spatial and temporal resolution of different tools and techniques in neuroscience. Time is on the *x*-axis and size is on the *y*-axis. (Adapted from Baars and Gage 2010)

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.

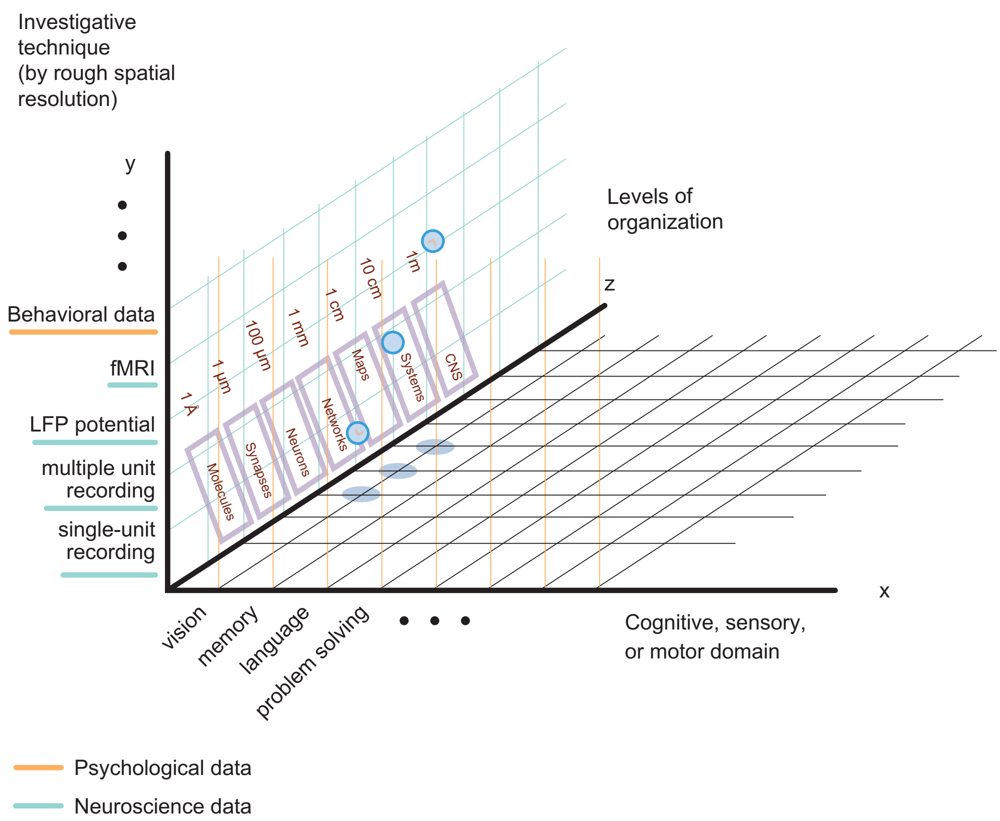


Figure 0.5 The "space" of contemporary cognitive science.

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

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

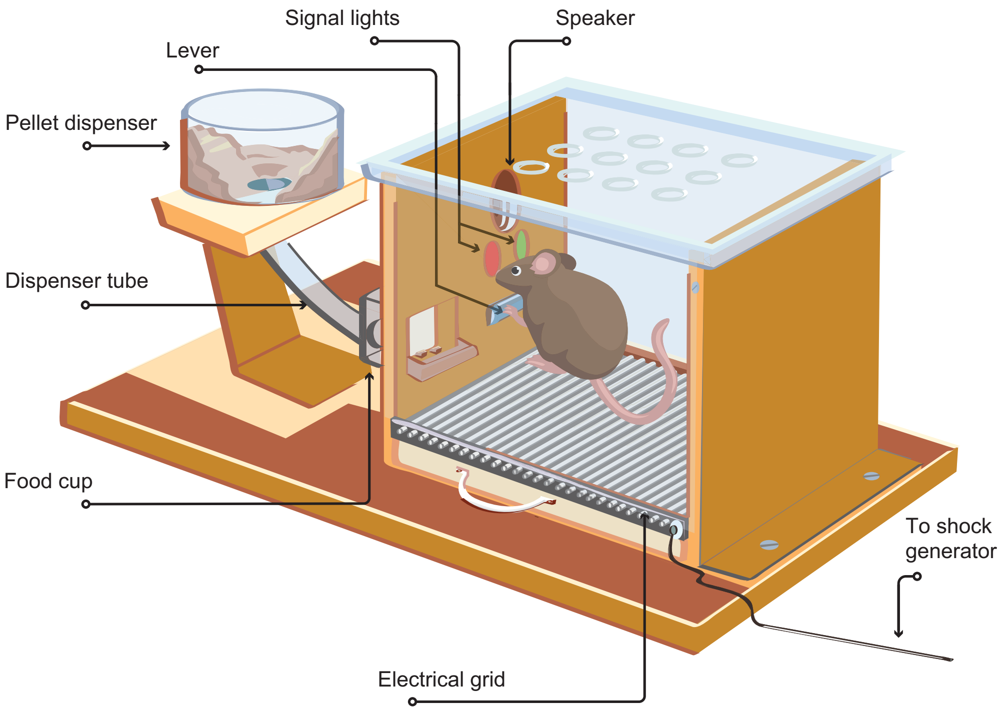


Figure 1.1 A rat in a Skinner box. The rat has a response lever controlling the delivery of food, as well as devices allowing different types of stimuli to be produced.

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

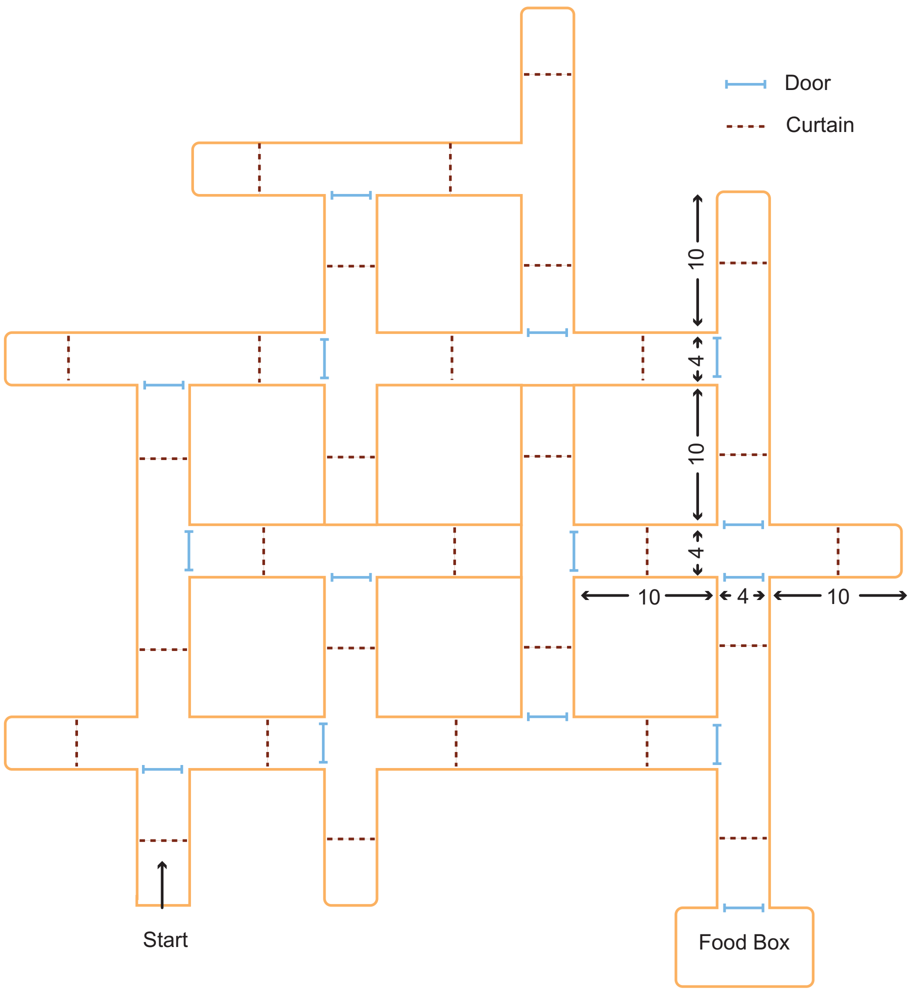


Figure 1.2 A fourteen-unit T-Alley maze (measurements in inches). Note the blocked passages and dead ends. (Adapted from Elliott 1928)

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.