

COGNITIVE SCIENCE

AN INTRODUCTION

second edition



Nell A. Stillings
Steven E. Welsler
Christopher H. Chase
Mark H. Feinstein
Jay L. Garfield
Edwina L. Rissland

Cognitive Science

An Introduction Second Edition

*Neil A. Stillings,
Steven E. Weisler,
Christopher H. Chase,
Mark H. Feinstein,
Jay L. Garfield, and
Edwina L. Rissland*

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England

© 1995 Massachusetts Institute of Technology

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from the publisher.

This book was set in Palatino by Asco Trade Typesetting Ltd., Hong Kong and was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Cognitive science: an introduction / Neil A. Stillings . . . [et al.].

—2nd ed.

p. cm.

"A Bradford book."

Includes bibliographical references.

ISBN 0-262-19353-1

1. Cognition. 2. Cognitive science. I. Stillings, Neil A.

BF311.C5523 1995

153—dc20

94-29553

CIP

Second printing, 1998

To our parents

Contents

List of Authors	xi
Preface to the Second Edition	xiii
Preface to the First Edition	xv
Note to the Teacher	xvii
Chapter 1 What Is Cognitive Science?	1
1.1 The Cognitive View	1
1.2 Some Fundamental Concepts	2
1.3 Information Processes Can Be Analyzed at Several Levels	7
1.4 Computers in Cognitive Science	11
1.5 Applied Cognitive Science	12
1.6 The Interdisciplinary Nature of Cognitive Science	13
Suggested Readings	14
References	14
Chapter 2 Cognitive Psychology: The Architecture of the Mind	15
2.1 The Nature of Cognitive Psychology	15
2.2 The Notion of Cognitive Architecture	16
2.3 A Global View of the Cognitive Architecture	18
2.4 Propositional Representation	26
2.5 Schematic Representation	32
2.6 Cognitive Processes, Working Memory, and Attention	37
2.7 Mental Images	42
2.8 Automatic and Controlled Processes	55
2.9 The Acquisition of Skill	58
2.10 The Connectionist Approach to Cognitive Architecture	63
Suggested Readings	83
References	83

Chapter 3 Cognitive Psychology: Further Explorations	87
3.1 Concepts and Categories	87
3.2 Memory	99
3.3 Reasoning	116
3.4 Problem Solving	129
Suggested Readings	135
References	136
Chapter 4 Artificial Intelligence: Knowledge Representation	139
4.1 The Nature of Artificial Intelligence	139
4.2 Knowledge Representation	151
Suggested Readings	173
References	174
Chapter 5 Artificial Intelligence: Search, Control, and Learning	177
5.1 Search and Control	177
5.2 Learning	192
Suggested Readings	212
References	213
Chapter 6 Linguistics: The Representation of Language	215
6.1 The Study of Linguistic Knowledge	215
6.2 Phonology	220
6.3 Syntax	241
6.4 Universals	260
Suggested Readings	267
References	267
Chapter 7 Neuroscience: Brain and Cognition	269
7.1 Introduction to the Study of the Nervous System	269
7.2 Organization of the Central Nervous System	270
7.3 Neural Representation	291
7.4 Neuropsychology	306
7.5 Computational Neuroscience	323
Suggested Readings	325
References	326

Chapter 8 Philosophy: Foundations of Cognitive Science	331
8.1 Philosophy in Cognitive Science	331
8.2 The Enterprise of Cognitive Science	335
8.3 Ontological Issues	345
8.4 Epistemological Issues	367
8.5 The State of Cognitive Science	373
Suggested Readings	374
References	375
Chapter 9 Language Acquisition	379
9.1 Milestones in Acquisition	380
9.2 Theoretical Perspectives	392
Suggested Readings	405
References	406
Chapter 10 Semantics	409
10.1 Semantics and Cognitive Science	409
10.2 Meaning and Entailment	410
10.3 Reference	411
10.4 Sense	418
10.5 Problems in Possible-Worlds Semantics	424
10.6 Cognitive and Computational Models of Semantic Processing	426
Suggested Readings	432
References	433
Chapter 11 Natural Language Processing	435
11.1 Preliminaries	435
11.2 On the Role of Grammar in Language Processing	436
11.3 Connectionist Models	445
11.4 On the Role of Discourse	447
11.5 More on the Role of General Knowledge	452
11.6 Production	454
11.7 Conclusion	458
Suggested Readings	458
References	458

Chapter 12 Vision	461
12.1 The Problem of Vision	461
12.2 Low-Level Visual Processes	464
12.3 Intermediate Processes and Representations in Vision	479
12.4 High-Level Visual Processes	490
12.5 The Architecture of Visual Computation	502
Suggested Readings	512
References	512
Index	517

List of Authors

Neil A. Stillings

Ph.D. in psychology, Stanford University. Professor of Psychology in the School of Communications and Cognitive Science, Hampshire College. A founding member of Hampshire's cognitive science program with over twenty years of teaching experience. Organizer of the 1986 national workshop on teaching cognitive science, funded by the Sloan Foundation, and the 1993 national workshop on undergraduate cognitive science, funded by the National Science Foundation. Coeditor of the entire textbook and primary author of chapters 1, 2, 3, and 12.

Steven E. Weisler

Ph.D. in linguistics, Stanford University. Associate Professor of Linguistics in the School of Communications and Cognitive Science, Hampshire College. Research and publications in syntax and semantics. Coorganizer of the 1993 national workshop on undergraduate cognitive science, funded by the National Science Foundation. Coeditor of the entire textbook, primary author of chapter 9, and coauthor of chapters 6 and 10.

Christopher H. Chase

Ph.D. in neuroscience, University of California at San Diego. Associate Professor of Cognitive Science in the School of Communications and Cognitive Science, Hampshire College. Research and publications in cognitive neuroscience, particularly reading development and developmental dyslexia. Primary author of chapter 7.

Mark H. Feinstein

Ph.D. in linguistics, City University of New York. Professor of Linguistics in the School of Communications and Cognitive Science, Hampshire College. Member of the Hampshire faculty since 1976. Research and publications in phonology, mammalian vocalization, animal cognition, and evolution of cognition and behavior. Primary author of chapter 11 and coauthor of chapter 6.

Jay L. Garfield

Ph.D. in philosophy, University of Pittsburgh. Professor of Philosophy in the School of Communications and Cognitive Science, Hampshire College. Research and publications in foundations of cognitive science, philosophy of mind, philosophy of language, metaphysics, epistemology, and Buddhist philosophy. Directs Hampshire's exchange program with the Tibetan universities in exile in India. Author of *Belief In Psychology* (MIT Press) and author or editor of other books. Primary author of chapter 8 and coauthor of chapter 10.

Edwina L. Rissland

Ph.D. in mathematics, MIT. Professor of Computer Science, University of Massachusetts at Amherst and Lecturer on Law, Harvard University Law School. Research and publications on knowledge representation, case-based reasoning, artificial intelligence and the law, and machine learning. Fellow of the American Association for Artificial Intelligence and Vice President of the International Association for Artificial Intelligence and Law. Primary author of chapters 4 and 5.

Preface to the Second Edition

The second edition of *Cognitive Science: An Introduction* retains the organizational format and the level of presentation of the first edition. Many chapters of the book have been substantially revised, however, and several new chapter sections have been added. The changes are detailed in the Note to the Teacher. The team of authors who wrote the first edition remains largely unchanged. David Rosenbaum and Lynne Baker-Ward have gone on to other pursuits. We have particularly missed David's colleague-ship and his ability to put his love of the field into words. Fortunately, Christopher Chase joined the Hampshire College faculty at just the right time and volunteered to take over the chapter on neuroscience.

One or two people were primarily responsible for each chapter: Stillings, chapters 1, 2, 3, 12, and several passages on neural computation in chapter 7; Rissland, chapters 4 and 5; Feinstein and Weisler, chapter 6; Chase, chapter 7; Garfield, chapter 8; Weisler, chapter 9; Garfield and Weisler, chapter 10; and Feinstein, chapter 11. Neil Stillings and Steven Weisler edited the entire manuscript.

Many people deserve our thanks for helping with the preparation of the second edition. The authors are part of the cognitive science community in the five-college area centered in Amherst, Massachusetts. It is impossible to mention all of the colleagues and students who contributed to our work, but we thank them for their help. We appreciate the suggestions and corrections of people who wrote to us about the first edition. Patricia Churchland, Richard Thompson, Lee Spector, Jamie Callan, Paul Utgoff, and Jody Daniels reviewed parts of the manuscript for the second edition and made many helpful suggestions. Conversations with Gary Marcus, Sean Stromsten, Sean Hill, Andy Barto, and Richard Yee were also very helpful. Rebecca Neimark prepared the new figures with admirable patience and professionalism. Michael Zenner, Michael Patterson, and Stacey Guess helped with references and with the computerized scanning of the first edition. Leni Bowen helped with permissions, typing, and editing at several crucial points, and she kept Hampshire College's cognitive science program on an even keel throughout our work.

We could not have seen this revision through without Betty Stanton's generosity, faith, and friendly persistence and Teri Mendelsohn's professionalism and unfailing good cheer. The copy editors, Anne Mark and David Anderson, caught our lapses in clarity and our stylistic imperfections with almost frightening accuracy, and they even corrected our antiquated Beethoven scholarship.

Katherine Pfister and Monica and Tim Stillings deserve special thanks for their support for Neil Stillings during the preparation of this edition.

Preface to the First Edition

The seven authors of this book have worked together for many years in the cognitive science community that has grown up in the five-college area surrounding Amherst, Massachusetts. Six of us (Stillings, Feinstein, Garfield, Rosenbaum, Weisler, and Baker-Ward) teach in Hampshire College's School of Communications and Cognitive Science (Baker-Ward recently moved to North Carolina State University). The seventh, Edwina Rissland, is a member of the Computer and Information Science Department at the University of Massachusetts. We have all also been members of the interdepartmental graduate program in cognitive science at the University of Massachusetts.

The cognitive science program at Hampshire College is a unique educational experiment. Hampshire was planned during the 1960s as an experimenting college, and cognitive science was one of the innovative programs it offered when it opened its doors in 1970. In 1972 a single academic department (then called the School of Language and Communication) was formed to house the cognitive science faculty along with faculty members in computer science and communications. The authors would like to thank the college and its leadership for allowing us to develop our commitment to the idea that cognitive science could be an exciting area of undergraduate study. We would also like to thank William Marsh, who was the co-founder and for many years the leader of the program.

Although we worked jointly to lay out the book and rework the manuscript, one or two people were primarily responsible for each chapter: Stillings, chapters 1, 2, 3, and 12; Rissland, chapters 4 and 5; Feinstein and Weisler, chapter 6; Rosenbaum, chapter 7; Garfield, chapter 8; Weisler, chapter 9; Garfield and Weisler, chapter 10; Feinstein, chapter 11; and Baker-Ward, the cognitive development section of chapter 3. Neil Stillings edited the entire manuscript.

Because seven people contributed to the book and because it is rooted in fifteen years of curricular experimentation and development at Hampshire College and the University of Massachusetts, it is impossible to mention everyone who contributed to its development. We thank all of the friends and colleagues who helped bring the book into existence.

The first draft of the book was written with the support of a faculty development grant to the authors from the Charles A. Dana foundation. The completion of the book would have been impossible without the enthusiastic support of Betty Stanton of The MIT Press.

The following people reviewed various parts of the manuscript at various points in its preparation. Gary Dell, Stevan Harnad, Keith Holyoak, and Zenon Pylyshyn deserve particular thanks for reading large chunks of the manuscript. Michael Arbib, Kevin Ashley, Robert Berwick, Charles Briggs, Carol Christensen, Charles Clifton, Bo Dahlbohm, Willem DeVries, Michael Gazzaniga, Allen Hanson, Norbert Hornstein,

John Haugeland, Haym Hirsch, David Kelley, Judith Kroll, David LeBeaux, Wendy Lehnert, Ken Livingston, Dan Lloyd, Lynn Nadel, Marty Ringle, Penni Sibun, Catherine Sophian, Devika Subramaniam, Paul Utgoff, Marie Vaughn, Robert Wall, Thomas Wasow, Sandy Waxman, and Bonnie Webber all read one or more chapters. All of these people exercised their professional expertise with care, and in addition many of them shared their thoughts on the nature of cognitive science and the challenge of teaching it to undergraduates.

We owe a great debt to Anne Mark, who served as manuscript editor. Her sense of expository and organizational clarity and her eye for both conceptual and typographical error are staggering. Nelda Jansen turned our ideas for figures into finished art work with efficiency and unfailing good humor. Erik Antelman and John Gunther also provided invaluable assistance in preparing the figures.

David Rosenbaum's preparation of chapter 7 was partially supported by a Research Career Development Award from the National Institute of Neurological and Communicative Disorders and Stroke (1 K04 NS00942-01) and by a grant from the National Science Foundation (BNS-8408634). The preparation of chapter 12 was partially supported by a grant from the National Science Foundation (SER-8163019) to Neil Stillings.

Ruth Hammen and Leni Bowen kept the School of Communications and Cognitive Science at Hampshire College running smoothly throughout the preparation of the book and lent a hand at several crucial points. The family members and friends of the authors helped with the special stresses of textbook writing. Katherine Pfister deserves special thanks for her steadfast support and good advice while Neil Stillings was editing the manuscript.

Note to the Teacher

We wrote *Cognitive Science: An Introduction* because we believe that undergraduate cognitive science programs ought to be offered to students and that an introductory course ought to be part of the cognitive science curriculum. At the undergraduate level, we feel that cognitive science is best conceived of as a broad interdisciplinary field that draws primarily on psychology, artificial intelligence, linguistics, philosophy, and neuroscience. The disciplines are to some extent distinct in their methods, theories, and results, yet they are strikingly unified by the convergence of their core questions and by the emergence in each of them of a computational, or information processing, view. In this text we try to maintain a consistent computational viewpoint, while honoring the distinctive contributions of each of the disciplines.

In our view the claim that cognitive science is distinguished by a computational or information-processing approach should not be taken too narrowly. The term *computational* should not be taken to mean that artificial intelligence ought to be the central or dominant discipline in cognitive science rather than an equal partner. The term *information processing* should not be restricted to a particular kind of process modeling that has been popular in cognitive psychology. As explained in chapters 1 and 2, we use these terms to refer to any research that concerns cognitive phenomena, such as perception, thought, or language, and that includes abstract levels of analysis that are designed to explain the functional significance of the inner workings of intelligent systems. In this book linguistic competence theories, production systems, parallel distributed processing models, the analysis of functional neural pathways in perceptual systems, and a wide range of other research is all considered computational.

Chapters 2 through 8 have explicit disciplinary origins. The reasons for maintaining some degree of disciplinary identity are both theoretical and practical. On the theoretical side, there is not yet a unified view of cognitive science that is widely accepted and that erases disciplinary boundaries. We feel that students deserve early exposure to the actual diversity and theoretical ferment in cognitive science. We have tried to write a book that is not a polemic for a premature and unstable unification of the field. On the practical side, the institutional structure of cognitive science continues to reflect its origins in the contributing disciplines. Most cognitive science programs are interdepartmental. Most faculty members hold degrees in one of the contributing disciplines. Many students gravitate toward cognitive science after an initial interest in one of the disciplines and go on to graduate study in disciplinary departments. For the time being, then, most teaching staffs and students are comfortable with an introductory course that contains some clear disciplinary signposts. Cognitive science groups should find it easy to build their locally developed emphases and syntheses into a course by emphasizing certain chapters in the text over others, stressing selected issues in lectures, and possibly including some supplementary reading.

The twelve chapters of the book reflect two useful themes for course planning. First, emphasizing the disciplines that contribute to cognitive science, the book consists of an introductory chapter, followed by seven chapters with disciplinary orientations, followed by four chapters on interdisciplinary research topics. Second, emphasizing the core topics of cognitive science, chapters 1 through 5 present a unified introduction to representation and thought, chapters 6, 9, 10, and 11 treat language, chapter 12 treats vision, and chapters 7 and 8 treat the neural and philosophical foundations of the field. A course that emphasized topics might cluster the chapters on language and also shift chapters 7, 8, or 12 out of order. The following remarks about the chapters can be used as an initial guide to course planning. The varying lengths of the chapters should be taken into account in deciding how much time to devote to them.

Instructors who are familiar with the first edition of the book should pay particular attention to the changes that have been made for the second edition. The new edition is about twenty percent longer than the previous one. Although the organizational format of the first edition has been retained, significant revisions, noted in the remarks below, have been made in most chapters.

1. "What Is Cognitive Science?" The concepts of representation and formal system are briefly introduced. The notion of levels of analysis is covered more thoroughly than in the first edition. The role of computers in cognitive science is discussed. The interdisciplinary nature of cognitive science is outlined (although the example analyzing the word *the* has been dropped from this edition). The material introduced in this chapter is expanded and reviewed throughout the book.
2. "Cognitive Psychology: The Architecture of the Mind" This chapter was substantially revised for the second edition. The classical view of the cognitive architecture, organized around the concept of physical symbol systems, is explicitly compared to the connectionist, or parallel distributed processing, view, of the cognitive architecture. The foundational material in chapter 1 is deepened considerably in sections 2.3 and 2.10, which are new to this edition. In the context of the classical view, propositional networks, schemas, working memory, imagery, and skill acquisition are introduced in some depth. The presentation is similar to the first edition. Section 2.10 introduces both the mechanics of connectionist networks and the controversy over the relation between connectionism and the classical view. Chapters 1 and 2 constitute an introduction to the foundations of cognitive science and should be covered in sequence at the beginning of the course.
3. "Cognitive Psychology: Further Explorations" In this chapter the classical and connectionist notions of cognitive architecture developed in chapter 2 are applied to the study of concepts, memory, reasoning, and problem solving. The treatment in the first edition has been substantially revised to reflect connectionist research and other new developments. The chapter can be taught out of order, and the sections within it are largely independent of each other. Unfortunately, the section on cognitive development had to be dropped from this edition to hold down the length of the book.
4. "Artificial Intelligence: Knowledge Representation" The chapter begins with several case studies of AI programs, which give beginning students a feel for AI. Several approaches to knowledge representation are then introduced and

compared, building on chapter 2. Semantic networks, frames and scripts, production rules, and formal logic are covered. The presentation is largely the same as that in the first edition.

5. "Artificial Intelligence: Search, Control, and Learning" This chapter continues the exploration of AI, beginning with an introduction to search and control. Search algorithms are covered in some detail in order to introduce beginning students to algorithmic thinking. Symbolic approaches to learning are treated in some detail, establishing a useful contrast with the material in section 2.10. The first five chapters can be taught as a unified introduction to central cognitive representations and processes.
6. "Linguistics: The Representation of Language" Linguistic competence theories are introduced as theories of linguistic representation. The section on phonology has been substantially revised to reflect recent developments. The material on syntax is largely unchanged, although the presentation has been strengthened in various ways. The final section on linguistic universals has been expanded and deepened. One or more of chapters 9, 10, and 11 can be read immediately following this chapter.
7. "Neuroscience: Brain and Cognition" This chapter has been entirely rewritten. The introductory material on neuroscience has been expanded, reflecting our view that it is increasingly important for cognitive science students to be familiar with basic concepts in neuroscience. We decided to put more emphasis on fundamental knowledge than on the very latest developments, although the treatment of neural representation and computation and of neuropsychology has been deepened considerably as well. The chapter should be assigned after chapters 1 and 2 and before chapter 12.
8. "Philosophy: Foundations of Cognitive Science" The chapter begins with a review and extension of the foundational concepts introduced in chapters 1 and 2 and goes on to cover philosophical issues concerning functionalism, propositional attitudes, qualia, and knowledge representation. The presentation is largely the same as that in the first edition, although several passages on philosophical issues raised by connectionism have been added. The chapter should be assigned after chapters 1 and 2.
9. "Language Acquisition" A description of the stages of language acquisition is followed by an introduction to the theoretical perspectives that have arisen out of linguistic innateness theories and recent formal work on learnability and parsing. The chapter has been revised and expanded to include more recent research.
10. "Semantics" This chapter is a relatively informal introduction to formal semantics as it is studied by linguists and philosophers. It begins with a treatment of quantifiers, names, tense, and scope. Possible worlds and their application to the analysis of necessity and propositional attitudes are then introduced. The chapter closes with a revised section on the role of formal semantics in psychology and AI. Although the chapter is self-contained, many of the basic ideas in it are introduced in chapters 1 through 4 and chapter 6.
11. "Natural Language Processing" The roles of grammar, discourse, and general knowledge in human and machine language understanding are discussed. The chapter also includes new sections on connectionist models and on language production.

12. "Vision" The format of this chapter is unchanged. Low-, intermediate-, and high-level vision are all covered in some depth. The perspectives of AI, psychology, and neuroscience are integrated. Although most of the material from the first edition has been retained, several recent lines of research have been added. The final section of the chapter on the architecture of visual computation has been completely rewritten to reflect new results in neuroscience and connectionism.

We believe that the introductory course in cognitive science should be offered to students in their first or second year of college-level study. The course will typically contain a mix of beginning students and older students who have some experience with one (or possibly more) of the contributing disciplines. We have tried to make the book accessible to committed first-year students. We have also included material that is challenging to these students and at the same time holds the interest of older, more sophisticated students. Lectures, discussions, and assignments should also be geared toward a range of intellectual sophistication, covering the fundamentals as well as pushing into more difficult material.

We also believe that an introductory course is a useful component of graduate programs in cognitive science. Because it provides a quick introduction to the multiple theories, results, and methods of cognitive science, this book should be useful in graduate courses, particularly when it is supplemented by more advanced readings.

Although the initial investment of time can be substantial, planning and teaching a cognitive science course can be an intellectually exciting experience that involves reflecting on the foundations of one's field, learning some new areas, and working closely with colleagues. The field of cognitive science depends on the strength of its introductory courses. We urge our colleagues around the world to make the necessary investment in undergraduate instruction.

Writing the second edition of this text has been an even more humbling experience than writing the first. Cognitive science has changed at a dazzling rate in recent years. Please write to the first author at the address below with your comments and suggestions. You can also request a copy of the National Science Foundation sponsored report on undergraduate cognitive science.

NEIL SILLINGS
NSILLINGS@HAMP.HAMPSHIRE.EDU

SCHOOL OF COMMUNICATIONS AND COGNITIVE SCIENCE
HAMPSHIRE COLLEGE
AMHERST, MA 01002

Chapter 1

What Is Cognitive Science?

One of the most important intellectual developments of the past few decades has been the birth of an exciting new interdisciplinary field called *cognitive science*. Researchers in psychology, linguistics, computer science, philosophy, and neuroscience realized that they were asking many of the same questions about the nature of the human mind and that they had developed complementary and potentially synergistic methods of investigation. The word *cognitive* refers to perceiving and knowing. Thus, cognitive science is the science of mind. Cognitive scientists seek to understand perceiving, thinking, remembering, understanding language, learning, and other mental phenomena. Their research is remarkably diverse. It includes, for example, observing children, programming computers to do complex problem solving, analyzing the nature of meaning, and studying the principles of neural circuitry in the brain.

1.1 The Cognitive View

Like all intellectual disciplines, cognitive science involves the adoption of a definite perspective. Cognitive scientists view the human mind as a complex system that receives, stores, retrieves, transforms, and transmits information. These operations on information are called *computations* or *information processes*, and the view of the mind is called the *computational* or *information-processing* view.

The perspective of cognitive science, although it is necessarily partial, provides a unique and rich set of insights into human nature and the human potential, including our potential to develop more powerful information technologies. The cognitive view arose from some quite natural puzzles that fascinate all of us at one time or another. One historical origin of cognitive science was the Greek philosophers' interest in deductive reasoning, the process by which one assumes some information to be true and derives further information that follows logically from the assumptions. For example, from the premises *All dogs have fleas* and *Fido is a dog*, one can logically derive the conclusion *Fido has fleas*. In his theory of syllogistic reasoning, Aristotle showed that deductively valid arguments often take one of a small number of general forms. Learning to reason deductively, then, can be viewed as learning an information process by which valid forms of argument can be recognized and produced. By the seventeenth century the philosophers Leibniz and Hobbes were arguing that all thought was a kind of calculation with nonnumerical information. Today, as we will see in the chapters that follow, linguists and philosophers continue to study the logical properties of language, cognitive psychologists compare people's actual reasoning processes to the idealized systems devised by philosophers and mathematicians, and researchers in artificial intelligence write computer programs to do logical reasoning.

To take another example, consider Louis Armstrong's brilliant improvised solo on his 1927 recording of "Potato Head Blues," which was a landmark in his decisive contribution to the development of jazz. We could adopt many different perspectives on this event. A physicist and biologist might collaborate to study how Armstrong shaped his notes in terms of the production of an air stream by certain complex muscle contractions and the physical properties of the vibrating trumpet. A musicologist might study how the African-American community in New Orleans combined properties of African and European music during the late nineteenth and early twentieth centuries. A music theorist might study how jazz improvisation is governed by the harmonic structure, or chord "changes," of a tune. But the music theorist's analysis raises fascinating questions about human information processing. How was Armstrong able to learn the rules and creative possibilities of harmony without any formal training in music theory? In his improvisation he was essentially composing music while he played. How could he possibly create compositions of great structural elegance so quickly? Part of the answer must be that the human mind has general characteristics that make it possible to learn the rules of improvisation and to apply them with great facility. Recent research on skill acquisition, described in chapter 2, has uncovered some of these characteristics. The results of the research could be used to help design new methods to teach musicians to improvise.

Many other examples could be given of the naturalness of the information-processing view. Whatever else people are doing, they are always taking in, storing, retrieving, transforming, transmitting, and acting on the basis of information. Our knowledge of the general characteristics of human information processing always provides one interesting perspective on human activity. This knowledge is equally relevant as we confront the intellectual potential of computers, wondering what their special strengths and limitations might be.

1.2 Some Fundamental Concepts

Understanding how the computational, or information-processing, view of the mind has led to a significant body of scientific research requires a more detailed understanding of what it means to say that something is a computational system. To set the stage for the following chapters, we will begin to explain and illustrate some fundamental concepts here. We will illustrate these concepts with the example of a familiar information process, doing arithmetic. Specifically, we will analyze the multiplication of nonnegative whole numbers.

Information Processes Are Contentful and Purposeful

The first concept is that an important part of understanding an information process is understanding its significance, or the purpose that it serves. An information process typically allows an organism or system to make systematic responses to some range of environmental conditions. The responses are typically adaptive or goal oriented. The information process has a quality of meaningfulness about it. That is, we think of the information in the system as being *about* the world, as having content, significance, or meaning. Such contentful qualities are also sometimes called *semantic*, or intentional, qualities. We find it natural to ascribe understanding, beliefs, or knowledge to any system that acts adaptively in response to information that it takes in from its environment. An understanding of an information process (or information-processing system)

will thus include an account of the content of the information it deals with as well as an account of its *competence* to employ that information in the service of certain goals.

Let us consider a simple, but instructive, example, whole-number multiplication. Without philosophical pause, we will stipulate that there is a world of numbers and relations among them that exists independently of any organism or device that carries out numerical computations. The product of two numbers is a third number that is determined by a function, or mapping, from pairs of numbers onto single numbers. This function, call it $\cdot$, is defined by the following characteristics, which must hold for all numbers x , y , and z : $x \cdot 0 = 0$; $x \cdot 1 = x$; $x \cdot y = y \cdot x$; $x \cdot (y \cdot z) = (x \cdot y) \cdot z$; and if $z > 0$ and $x > y$, then $x \cdot z > y \cdot z$. Also, the multiplication and addition functions are related. If we call addition $+$, then for all x , y , and z , $x \cdot (y + z) = x \cdot y + x \cdot z$.

For a device to qualify as a multiplier it must possess *representations* or *symbols* that stand for numbers. It must be able to accept and produce these symbols, and it must be able to transform them in a way that faithfully represents the product function as it was just defined. That is, when the device is fed two symbols, representing two numbers, it must produce the symbol that represents the product of the two numbers. Notice that by focusing on the purpose of a process, on the competence that it has to display, we have achieved a certain kind of purity. We have specified what a multiplier must do (its response function or input-output behavior) without specifying *how* it might do it, beyond requiring that it possess some kind of representation of numbers.

An analysis concerned with the structure of a system's environment, with the information that the system has at its disposal, and with the goals it can satisfy through its deployment of that information, can be called a semantic analysis, a competence theory, a knowledge-level analysis, or an ecological theory (if there is a special emphasis on an analysis of the environment and adaptiveness to it). We all use this kind of explanation in our everyday lives. We make assumptions about what people know and what their goals are, and we assume that they can, for the most part, use their knowledge rationally to meet their goals. This *folk psychology* is quite successful. It is a crucial part of our lives that we are able to predict and explain our own and others' behavior at the knowledge level much of the time.

Given our conscious awareness of our own beliefs and goals, and our everyday methods of discovering other people's, it might be imagined that analyzing the human mind at the knowledge level is not a significant scientific problem. However, cognitive scientists have found that in several important areas our awareness of the information we possess is limited and even misleading. An example of the limitations of awareness is our judgment of the distance from us of an object in our visual field. We know that we possess depth perception. In scientific research on vision, however, it becomes important to know just how accurate depth judgments are, whether they are affected by the familiarity of the object, whether they depend on the presence of other objects in the scene or on the visibility of the ground, and so on. We cannot answer such questions just by consulting our visual awareness. Careful laboratory studies as well as careful studies of the visual environment are required. By studying the environment, we might find that the presence of other objects in a scene can provide information about depth. For example, a nearer object can partly occlude the visibility of a more distant object because it blocks the light reflected from part of that object. Occlusion is thus a possible cue to depth. We could then study in the laboratory whether our depth judgments make use of occlusion information, and, if so, how that information interacts with other information about depth.

An example of how awareness can be misleading is our feeling that we have a uniformly vivid impression of the visual world that is nearly 180 degrees wide. It is rather easy to demonstrate that this impression is an illusion and that we are not able, for example, to perceive much detail in the periphery of the visual field without moving our eyes. Given that our awareness is a poor guide to the information that is immediately available in different parts of the visual field, careful laboratory research is again the only route to discovering our visual competence. In line with these examples from vision, it will be shown in chapter 6 that we are largely unaware of our competence with the sound and grammatical structures of our native languages.

If, rather than trying to understand an existing system, we are trying to *construct* a system to carry out a process, it is obviously also very helpful to begin with a clear understanding of what we are trying to accomplish. It would be unfortunate, for example, if the designers of an electronic calculator did not fully understand the correct behavior of the arithmetic functions that their machine was supposed to calculate. For this reason, a surprising amount of the research in artificial intelligence involves trying to figure out just what knowledge is involved in certain types of intelligent behavior.

Information Processes are Representational

To characterize an analysis at the knowledge level, we have already invoked our second key concept, *representation*. The information that figures in a computation must be represented in some way. Further understanding of a computation requires an understanding of how the information is represented. Returning to the example of multiplication, we have so far said that a multiplier must represent numbers, but we have not said how. To develop more detail, let us suppose that the representation is the familiar decimal place-value notation. In this notation there are ten digit symbols, "0," "1," "2," . . . "9," which represent the numbers 0 through 9. Notice that we distinguish here between symbols, such as "5," which are part of the representation, and numbers, such as 5, which are being represented, by placing quotation marks around the symbols. The distinction can be brought out more clearly by recalling the Roman notation system, in which "V" is the symbol for 5, there is no single digit symbol for the number 2, there is a single digit symbol for the number 50, namely, "L," and so on. In decimal notation numbers larger than 9 are represented by *concatenating* the digit symbols into *strings*, such as "65," which is a string of length 2 with "5" in the first position and "6" in the second position. The assignment of a number to a symbolic string is determined systematically by the decimal place-value function, in which the contribution of each digit symbol to the assignment is determined jointly by its basic assignment and its position in the string. For example, knowing that the basic assignments of "6" and "5" are 6 and 5 and knowing the place-value function, one can determine that the assignment of "65" is $6 \cdot 10^1 + 5 \cdot 10^0$, which equals $60 + 5$, or 65. The decimal place-value function establishes a one-to-one correspondence, or mapping, between digit strings and numbers. Note that although the Roman system also uses strings of digit symbols to represent numbers, the mapping between strings and numbers cannot be defined as a place-value function (consider, for example, the significance of the digit symbol "I" in the left position of the strings "II" and "IV").

Place-value representation illustrates some properties that can give representations generality and power. First, a potentially infinite set of symbols can be constructed by beginning with a relatively small stock of basic symbols and using one or more rules of construction, or *syntax*, to build new, complex symbols. In the case of whole num-

bers the ten digit symbols are the basic symbols, and concatenation is the sole syntactic rule. Allowing strings of digits to be of unlimited length produces a potentially infinite stock of symbols. Representational schemes that include rules for building complex symbolic structures out of simpler ones are called *combinatorial*, *generative*, or *productive*. Second, the meaning, or *semantic interpretation*, of a complex symbol is built up from the meanings of the syntactic parts of the symbol. Typically, each basic symbol has a fixed meaning, and each syntactic rule is associated with a semantic rule that contributes to the interpretation of the complex symbol. For whole numbers, the numbers 0 through 9 are the fixed interpretations of the ten basic digit symbols "0" through "9." The syntactic rule of concatenation is associated with the place-value rule of interpretation. For example, if we already have a digit string "xyz," and we concatenate "5" onto it to make "5xyz," then the place-value rule says that the numerical interpretation of "5xyz" is $5 \cdot 10^3$ (or 5,000) plus the numerical interpretation of "xyz." Representational schemes in which the interpretation of complex symbolic structures is determined by the interpretations of their syntactic parts are said to have a *compositional semantics*. Third, information processes that transform symbolic input structures into symbolic outputs can be defined in terms of syntactic structures of the inputs and outputs. Such information processes analyze the syntactic structures of inputs and build syntactically structured outputs. Such information processes are also called *algorithms*. Because the meanings of symbolic structures are a function of their syntax, information processes that operate on syntax can produce meaningful results and represent meaningful operations in the domain that is represented.

These properties of algorithms can be illustrated by the example of multiplication. We have to define a process that operates syntactically on decimal place-value representation in such a way that the process succeeds in representing the multiplication function, $\cdot$. The well-known process for multiplying numbers with paper and pencil, call it $\cdot_{\text{pencil}}$, will serve our purposes. It is illustrated in the following example:

```

  65
 32
---
130
195
---
2080

```

The decimal multiplication algorithm is a fairly complicated information process. It requires, among other things, a memory for the products of all of the pairs of digits (e.g., "2" $\cdot$ "3" yields "6"), an ability to process columns in concatenation order starting with the "ones" column, an ability to "carry," an ability to save partial results until all columns are processed, and an ability to carry out paper-and-pencil decimal addition.

As we said in the previous section, the algorithm can be said to represent the product function if it operates in such a way that the representational mapping between symbols and numbers is maintained. Let us explore this in more detail by looking at the example above. The representation assigns the string "65" to the number 65, "32" to 32, and "2,080" to 2,080. The product of 65 and 32 is 2080. Therefore, when the longhand decimal multiplication algorithm is applied to "65" and "32," the output must be "2,080." More generally, if we map any two input numbers onto digit strings, apply the algorithm, and map the resulting string back to a number, that number must be the product of the input numbers. If the algorithm always mimics

the operation in the world, we say that they have the same structure, or are *isomorphic*. Notice, again, that it is critical to distinguish between the operation on symbols and the operation in the domain of numbers that it represents. Many different algorithms could be used to represent multiplication. For example, repeated addition is a possible algorithm for multiplication. To multiply "65" by "32" using this approach, we would apply an algorithm for addition to 32 copies of "65."

Information Processes Can be Described Formally

The third basic concept results from further attention to the notion of algorithm that was just developed. An algorithm is defined completely in terms of processes that operate on a representation. The processes do not operate on the domain being represented. They are not even defined in terms of the meaning of the representation, which is carried separately by the semantic mapping from the representation to the domain. An algorithm is a *formal* procedure or system, because it is defined in terms of the form of the representation rather than its meaning. It is purely a matter of manipulating patterns in the representation.

The algorithm for paper-and-pencil decimal multiplication, for example, is entirely a matter of manipulating strings of digit symbols according to certain rules. A person with no knowledge that the strings stand for numbers could be taught to carry out the rules correctly. In fact, most of us probably carry out these rules without thinking about why they lead to correct results. Why, for example, does carrying the 1 when multiplying 65 by 2 lead to a correct result? Exactly why is the 195 placed one space to the left before adding it to the 130? Answering these questions requires referring to the place-value mapping that determines the meaning of decimal notation. But one need not know about the place-value mapping to carry out the process. The process is meaningful, but it acquires its meaning indirectly.

An appreciation of the formal nature of information processes brings out two further points. First, since algorithms can be carried out without any higher knowledge about their meanings, they can be carried out by physical systems, which can be biological or engineered. This insight is one of the intellectual foundations for computer science and for the belief that a scientific understanding of mind can be achieved. Second, a formal analysis of an information process provides a particularly convincing demonstration that we understand its inner workings. Because it makes no reference to the meaning of a process, an algorithmic analysis of a process shows exactly how it is done without leaning on our understanding of its meaning.

In summary, the cognitive scientist approaches information processes by distinguishing between formal operations on symbols and the representational relations between symbols and what they stand for. The realization that an organism or machine can produce meaningful behavior by performing formal operations on symbolic structures that bear a representational relationship to the world is a key insight of cognitive science.

Our choice of example should not be taken to mean that cognitive science is easy. Cases like the multiplication algorithm or typical computer programs are good illustrations, because they show that the strategy of cognitive science makes sense. A computer program that multiplies numbers or does database management is obviously performing formal operations on information that has a well-defined representational relationship to a particular domain. Just as obviously, the success of the programs is

due entirely to the fact that the formal operations preserve the representational relationship. No inaccessible form of intelligence or understanding is present. Things are not so simple with biological organisms, however. The human mind was not built by a team of engineers and programmers who can just hand over the blueprints and programs. The information processes and representational relationships required for human intelligence are extremely complex, and they can only be discovered by careful and creative research.

Cognitive Science Is a Basic Science

A fourth point about the information-processing view is that the concepts just introduced are employed by cognitive scientists in the pursuit of basic scientific knowledge. Cognitive scientists seek to discover highly general and explanatory fundamental principles of information processing. This goal collides with the tremendous variability that we see in human thought and behavior. We might doubt that any principles of human information processing can be found that hold across all cultures and historical epochs. Certainly we will have to dig deeper than our simple example of the decimal multiplication algorithm, which is a relatively recent human invention that is spottily distributed among the world's people and may slowly disappear from human cognition as electronic calculators become more common. To take another example that is explored in detail in chapters 6 and 9, although people seem to communicate by making noises with their vocal tracts in all of the world's societies, languages show tremendous variation. Do the formal structures and processes of all of the world's languages have anything in common? Research suggests that important *linguistic universals* do indeed exist and that they play an important role in children's acquisition of their native languages. Of course, the hypothesized universals must also explain how the great surface variety in the world's languages is possible.

1.3 Information Processes Can Be Analyzed At Several Levels

The distinction between studying the competence or knowledge of a system and studying its formal information processes can be thought of as a distinction between levels of analysis. The formal analysis is at a *lower* level, providing an account of the information processes that *underlie* the competence that is visible at a *higher*, behavioral level. The analysis of the semantic mapping from the formal representations to the domain can be thought of as a bridge between the formal and knowledge levels. It explains why a formal system is a successful implementation of a particular competence. Although the formal analysis can be thought of as deeper than a knowledge-level analysis, it does not replace it. Each level of analysis contributes its own insights to the overall picture. Without the knowledge-level analysis, including the understanding of the representational mapping, we wouldn't have an understanding of what the algorithm accomplishes, and we wouldn't be able to capture the fact that two different algorithms with different representations both compute the same function. Without the formal analysis we would know what a system does but not how it does it. Reviewing the example of decimal multiplication, the abstract competence analysis ($x \cdot 0 = 0$, etc.) tells us what the product function is, the representational analysis shows that decimal notation systematically represents numbers, and the formal analysis fully specifies the mechanics of the algorithm. If we ignore a level of analysis, we miss an important part of the picture.

The use of multiple levels of analysis to study intelligent systems is one of the hallmarks of cognitive science. Research on a topic can be focused on one or another level of analysis at a given time, depending on the most promising avenues of study, and findings at one level can influence understanding at other levels. The goal is to develop an ever more insightful and complete picture of cognition at all levels of analysis.

Information Processes Must Be Physically Implemented

The analysis of systems at the knowledge level and the formal level, and the use of the concept of representation to tie these levels together, distinguish cognitive science (including computer science) from other sciences. To a surprising extent information processes can be studied at the knowledge and formal levels independently of their physical implementations. This is somewhat remarkable because an information process cannot actually occur unless it is implemented in some physical medium. For any actual information-processing system there obviously are physical levels of analysis aimed at uncovering its biology or physics. Further, there must be some relationship between the physical and formal levels of analysis, because the ability of a physical system to carry out a formal information process depends completely on its physical construction. Just as biologists concluded in the nineteenth century that life arises from particular organizations of matter and energy and not from a special life force, so cognitive scientists proceed from the assumption that cognition arises from material structure and processes and not from any mysterious extraphysical powers. These considerations raise the question of what role physical levels of analysis should play in cognitive science.

The answer to the question must proceed from an understanding of the relationship between the formal and physical levels. Just as a representational mapping ties the knowledge and formal levels of analysis together, so the formal and physical levels are tied together by an *implementational* mapping. The formally defined representations and processes must be mapped onto arrangements of some physical medium and transformations of that medium. The mapping must establish an isomorphism between the formal and physical levels, so that the states and changes in the physical medium faithfully preserve all of the informational relationships in the formal system. To return to our multiplication example, any physical system that implements the standard decimal multiplication algorithm must bear a precise correspondence to the algorithm. Consider the state of having a multiplicand of "65." A human being with pencil and paper physically implements this state by inscribing certain two-dimensional shapes in a particular left-to-right order in a particular position on the paper. In an electronic calculator a multiplicand of "65" might be implemented by the distribution of electrical charge in certain circuit components. In both cases there is a precise correspondence between physical states and transformations in the computing system and the formal structure of the algorithm.

The Necessity of Higher Levels of Analysis

We might imagine that our ultimate goal is to understand the physical workings of an information-processing system and that having a physical understanding would render the understanding at the formal and knowledge levels unnecessary. Consider an electronic calculator that can multiply as an example. There is a sense in which its behavior can be completely understood in terms of the movement of electrical current through

circuits made up of conductors and semiconductors. The physical analysis would make use of the circuit diagram, laws governing electrical circuits (e.g., Ohm's and Kirchhoff's laws), and perhaps some more detailed electron physics needed to understand the behavior of the semiconductors.

The most immediate problem with sticking with the physical analysis is that, by dispensing with the knowledge level, we lose any account of the significance of the calculator's behavior. The physical description does not capture the fact that the calculator is a multiplier. This loss of meaning has many consequences. A mistake in the circuit design, for example, would not be reflected in the physical description, since the circuit laws give equally nice descriptions of useful and useless circuits. Further, at the physical level we could not express the fact that two calculators of different design both multiply, because their physical descriptions could be arbitrarily different (e.g., an electronic calculator and an old-fashioned mechanical adding machine). This kind of inadequacy is generally referred to as the failure of a level of analysis to capture a generalization. The knowledge level is needed to capture generalizations that involve the significance or purpose of an information process or the content of its representations. The physical level of analysis obviously captures other generalizations that cannot be expressed at the knowledge level, such as Ohm's law.

Acknowledging the need for the knowledge level, we might still hope to dispense with the formal level of description. Given that we have a representational mapping from the world to the formal system and an implementational mapping from the formal system to a physical medium, we could imagine bypassing the formal level and working with a single direct map from the world to the physical medium. For example, we could map the world of numbers and the product function directly onto charge distributions and their fluctuations in the circuits of an electronic calculator without bothering with the formal analysis of place-value notation and the multiplication algorithm defined over it. Once again, however, the formal level captures generalizations that cannot be expressed on either the knowledge level or the physical level. In this case the generalizations have to do with the algorithms a system is using. Physical descriptions are full of physical detail that is irrelevant to the algorithm in question, and, worse, they cannot express the fact that two systems of different physical construction compute the same algorithm (e.g., a person and an electronic calculator both doing decimal multiplication). Algorithmic concerns also cannot be expressed at the knowledge level, because the knowledge level abstracts away from the formal detail of computation in order to lump together processes that have the same semantics in spite of having different algorithms. The generalizations that can be captured at the formal level are profound, forming a good part of the discipline of computer science. We will see in section 5.1, for good example, that it is possible to define and compare different algorithms for searching through large sets of alternatives without any consideration of either the semantics or the physical implementation of the alternatives.

Each level of analysis, then, has its own focus and descriptive vocabulary, allowing a conceptually clear explanation of some aspect of computational systems that is obscured or even lost at other levels of description. The concepts at each level express different generalizations and impose different categorizations on information-processing systems.

In addition to this conceptual reason for maintaining higher levels of analysis, there are methodological reasons as well. It is obvious that the design of engineered computational systems can be profitably guided by high-level decisions about what sorts of

algorithms they must compute and what parts of the world they will have to deal with. It is perhaps not so obvious, but equally important, that our attempts to understand biological information-processing systems, such as the human mind/brain, at the physical level can be guided by understanding at the knowledge and formal levels. Physically, the brain is enormously complex. We can get a much better idea of what to look at in the brain and how to look at it if we have a good idea of what the brain is computing and what algorithms it is using to do the computations. In particular, a theory of some mental process at the formal level is a source of hypotheses about the implementational mapping from formal representations and processes to nervous tissue. We will see in chapter 7, for example, that analysis of auditory localization in the barn owl at the behavioral and formal levels guided a systematic and successful search for the neural circuits that compute the direction of an environmental sound from the acoustic input to the ears.

The Importance of the Physical Level of Analysis

The fact that the knowledge and formal levels capture unique generalizations and that a particular formal system can be physically implemented in any number of ways suggests the possibility that cognitive science could be pursued without regard to the physical level of analysis, which could be left to biologists and engineers. Although it is true that the higher levels of analysis are the distinctive province of cognitive science and that much research in the field takes place exclusively at these levels, the physical level cannot be ignored, largely for practical reasons.

Understanding the physical implementation of information processes is obviously a crucial piece of an overall understanding of cognition, since all cognition occurs in some physical medium. Researching mappings between the higher levels of analysis and the physical level is greatly facilitated by an understanding of both levels. It is therefore not a good idea to enforce a complete division of scientific labor between people who work at the knowledge and formal levels and people who work at the physical level.

In general, results at one level of analysis can always potentially affect research at other levels, because the levels are mutually constraining. We have already pointed out that a formal theory of an information process constrains its physical implementation, because the physical implementation has to be a faithful instantiation of the formal process. The constraints run in the other direction as well. Discoveries about the physical structure of the brain can strongly suggest that it is computing certain kinds of algorithms and not others or that it represents knowledge in certain ways and not others. We will see in chapter 2, for example, that in recent years relatively simple facts about the brain, such as the speed with which neural cells can generate signals and the way in which the cells are interconnected, have been used by some researchers to draw far-reaching conclusions about the representations and algorithms that characterize cognition in biological organisms. In the realm of technology the availability of a particular kind of hardware can spur research into new representations and algorithms that are particularly suited to that hardware. For example, in recent years it has become feasible to build computers that have thousands of interconnected, active processors rather than just one. This possibility triggered a wave of research on parallel algorithms that could take advantage of many simultaneously active processors.

Ideally, research on cognition should occur at all levels of analysis, and all fruitful interactions among levels should be encouraged. Building bridges between the higher

levels of analysis and the physical level has been, and continues to be, the most difficult problem in the cognitive and neural sciences. In recent years, however, it has also been one of the most active areas of new research.

1.4 Computers In Cognitive Science

It should be clear from the discussion so far that a number of concepts and distinctions that are important in computer science are part of the underlying assumptions of cognitive science. More generally, the growth of computer science has greatly accelerated the development of cognitive science. Computers thus have a tendency to become a metaphor for all information processing. The metaphor is often useful, but it is potentially misleading as well.

The rough outline of the analogy between computers and the human mind should already be apparent. Our stress on the independence of an information process from its physical implementation is akin to the common distinction between software and hardware in the computer world. For example, when we see the same word-processing program running on two different brands of computers, it is obvious that an information process can be rigorously defined as a formal process without reference to any particular physical device. Current computer software also makes it clear that formal information processes can be powerful and flexible. Highly complex information, ranging from documents to personnel records to satellite photographs, can be represented and stored in computers. Programming languages allow finite sets of simple instructions to be flexibly combined to define complex operations on the stored information. Today's computers and computer software are concrete examples of information processors and processes that make it easier to understand what cognitive science is about.

On the other hand, computers are very far from exhibiting some of the most impressive, yet mundane, aspects of human cognition, such as learning a language, identifying the objects in a naturally occurring visual scene, or solving simple problems via analogies with other situations. Therefore, it is misleading to take the computer metaphor to mean that current computer programs, programming languages, and hardware are good models for human cognition. The usefulness for cognitive science of any current piece of computer science or technology is a matter of research, not prior stipulation. Research is now under way on new "generations" of computer software and hardware that many workers in the field think will be needed to advance computer intelligence.

This research includes the part of computer science called *artificial intelligence*, or *AI*, which overlaps considerably with cognitive science. Many researchers in AI try to model their computer programs after human intelligence, and they derive inspiration from insights into human information processing that come from other disciplines in cognitive science, such as psychology and linguistics. The insights also flow in the other direction. Attempts by AI researchers to program systems that can understand language, see, or solve problems have led to new, testable hypotheses about human cognition. In its interplay with the other cognitive science disciplines AI provides a powerful alternative methodology for exploring and testing theories of cognition that supplements the empirical methods of psychology and linguistics. Chapters 4 and 5 of this book are about AI.

The technique of expressing a cognitive theory as a computer program and then running the program to explore the ramifications of the theory is now an important

tool throughout cognitive science. Many theories are complicated enough that it is impossible to figure out what their predictions are without simulating them on computers. There are differences of emphasis between AI and the computer simulation of theories of human cognition. Much AI research is oriented toward practical results. A program that is too slow, for example, might be of no use. On the other hand, a useful AI program might not use the same algorithms as a human being solving the same problem. Researchers oriented toward simulation have complementary goals. Their measure of success is whether their programs work the same way as human (or animal) cognitive processes. It doesn't matter to them if the computer simulation is impractically slow, or slower than the biological process it simulates, as long as it makes the right predictions. In spite of these differences there is no absolute line between AI and simulation research. Any AI program might turn out to be a good theory of human cognition, and any simulation program might prove to have practical applications.

1.5 Applied Cognitive Science

Cognitive scientific theories are leading increasingly to practical applications. Many of the applications are in the domain of education and learning. As cognitive psychologists have come to understand the information processes involved in reading, for example, they have begun to develop new ways of diagnosing and treating children's reading difficulties. Linguistic theory has led to much more precise knowledge of the speech impairments that follow strokes in the left hemisphere of the brain. This new knowledge promises to be useful in designing appropriate speech therapy and computer-driven linguistic prostheses for stroke victims.

Other applications of cognitive science are more surprising. Contemporary theories of human memory have been applied to the question of the reliability of legal witnesses. This research has already led to important changes in the role of eyewitness testimony in the legal process. Another example is the application of new psychological theories of skill acquisition and visual-motor imagination to the design of training programs for athletes.

Artificial intelligence research is leading to applications in expert systems and robotics. Expert systems have been developed to aid in configuring computer systems, exploring for oil, establishing the structure of complex organic compounds, and diagnosing diseases.

These developing applications are exciting, particularly to those within the field or thinking of going into it. The excitement should be tempered, however, by an awareness of the moral and political questions raised by applied cognitive science. Just as physicists must confront the implications of nuclear technology, and biologists the implications of genetic engineering, so cognitive scientists must confront the implications of knowledge technology. The potentially controversial applications of cognitive science research range from the possible development of a new generation of intelligence tests, which might be misused, to the large-scale introduction of intelligent robots in manufacturing industries, which might cause a massive loss or displacement of jobs. As in other sciences, the less and the more controversial applications often flow from the same underlying theoretical research. For example, results in computer vision might be used to design either a visual prosthesis for the blind or the control system of a cruise missile carrying a nuclear warhead. We hope that this book will provide the basic understanding of cognitive scientific theory that is needed to think about the

policy issues posed by new information-processing technologies. We also strongly recommend the study of relevant aspects of history, social science, and the humanities, because the perspective of cognitive science, although crucial, must be supplemented by other perspectives.

1.6 The Interdisciplinary Nature of Cognitive Science

The Five Disciplines

As noted earlier, cognitive science is an interdisciplinary field that has arisen from the convergence on a common set of questions by psychology, linguistics, computer science, philosophy, and neuroscience. The five contributing disciplines will undoubtedly retain their separate identities, because each of them involves a much larger set of concerns than the focus on a basic science of cognition. A more interesting question is whether cognitive science will become a distinct academic discipline in its own right, within which the contributions of the five converging disciplines become so thoroughly intermingled and transformed that they are no longer identifiable. This book presents cognitive science in its current form. The distinctive contributions of each of the five disciplines are highlighted, particularly in chapters 2 through 8. The topics selected for discussion, however, are those where the disciplines have shared the most common ground. Chapters 9 through 12 present several examples of research areas in which the interdisciplinary collaboration has been particularly close and has begun to obscure some of the differences among disciplines.

Obviously, no one of the five contributing disciplines encompasses the entire subject matter of cognitive science, and each discipline brings to the field a focus on particular areas. Neuroscientists are primarily concerned with the organization of the nervous system. Linguists are concerned with the structure of human language and the nature of language acquisition. Philosophers are concerned with logic and meaning, and with clarifying the fundamental concepts of cognitive science, such as information and knowledge. Psychologists are concerned with general human mental capacities, such as attention and memory. Computer scientists are concerned with the possibilities for AI. Cognitive science encompasses all of these concerns. Cognitive scientists, although they usually specialize in one or two of the contributing disciplines, benefit greatly from the cross-fertilization of all of them.

The most important differences among the five disciplines are in the research methods that they use to address the nature of mind. Psychologists emphasize controlled laboratory experiments and detailed, systematic observations of naturally occurring behaviors. Linguists test hypotheses about grammatical structure by analyzing speakers' intuitions about grammatical and ungrammatical sentences or by observing children's errors in speech. Researchers in AI test their theories by writing programs that exhibit intelligent behavior and observing where they break down. Philosophers probe the conceptual coherence of cognitive scientific theories and formulate general constraints that good theories must satisfy. Neuroscientists study the physiological basis of information processing in the brain.

The Study and Practice of Cognitive Science

From the student's point of view one of the main attractions of cognitive science is the diversity of its methods. The field accommodates a wide variety of personal intellectual

styles and preferences. Those who are attracted to laboratory work can pursue experimental psychology or neuroscience. Those who dislike laboratory work but who like abstract problems and careful logical analysis can pursue philosophy. Those who love the challenge of writing computer programs can pursue AI, and those who love thinking about language can pursue linguistics. Those who are fascinated by the development of children can pursue cognitive development. This list could go on, but the general point is that cognitive science needs researchers who are motivated by many different kinds of curiosity and who like to do many different kinds of work. The field thrives on the presence of people who ask the widest possible variety of questions about the mind.

Your own intellectual tastes will probably lead you to prefer some of the chapters in this book over others. It is natural to pick a specialty, but we urge you not to let your preferences lead you to neglect other chapters. Cognitive science depends on a continuing dialogue among its various specialties, and the intelligence of the dialogue requires people with a good grounding in the fundamentals of the contributing disciplines. Today, a new generation of cognitive scientists is learning these fundamentals as undergraduates and beginning graduate students. The resulting increase in mutual understanding will enhance the collaborative research of the future.

Suggested Readings

The material in this chapter is covered from a similar point of view and at a somewhat more advanced level in chapter 2 of Newell's *Unified Theories of Cognition* (1990). An advanced treatment is given in the first four chapters of Pylyshyn's *Computation and Cognition: Toward a Foundation for Cognitive Science* (1984). Chapters 8 and 9 of Churchland's *Neurophilosophy: Toward a Unified Science of the Mind/Brain* (1986) present a case for the cooperative pursuit of research at different levels of analysis as well as a case for giving neuroscience the guiding role. Gardner's *The Mind's New Science: A History of the Cognitive Revolution* (1985) presents a history of cognitive science.

References

- Churchland, P. S. (1986). *Neurophilosophy: Toward a unified science of the mind/brain*. Cambridge, Mass.: MIT Press.
- Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, Mass.: Harvard University Press.
- Pylyshyn, Z. W. (1984). *Computation and cognition: Toward a foundation for cognitive science*. Cambridge, Mass.: MIT Press.

Chapter 2

Cognitive Psychology: The Architecture of the Mind

2.1 The Nature of Cognitive Psychology

Psychologists formulate and test theories about the human mind and behavior. Cognitive psychology concerns human cognition, our capacities for sensory perception, memory, thinking, problem solving, and learning. The cognitive psychologist who is oriented toward cognitive science views the human mind as a remarkable information-processing system that is extraordinarily powerful in most circumstances and yet surprisingly limited in others. As you read this paragraph, the meaning of each word is effortlessly activated within a mental dictionary of tens of thousands of words. But you probably do not remember the names of the six coauthors of this book. A book-reading computer, using current technology, would have a much easier time remembering names than deploying meanings.

We begin this book with cognitive psychology for two reasons. First, cognitive psychology focuses on the human mind. In order to fully appreciate work in artificial intelligence (AI), it is necessary to have some familiarity with theories of human intelligence. Second, in keeping with the emphasis on basic science within cognitive science, cognitive psychologists have tried to develop theories of highly general cognitive capacities. They have asked what sort of general information-processing capacities a mind must have in order to do the many things it does. That is, they have tried to figure out what the overall design, or architecture, of the mind is. It is a good idea to begin the study of cognitive science by confronting some of the most basic questions about how it is possible for the mind to work as it does.

In order to maintain the focus on the most general cognitive capacities, we will delay discussion of two of the more specialized information-processing subsystems of the mind. Language will be extensively discussed in chapters 6, 9, 10, and 11. Vision will be discussed in chapter 12. Linguistic and visual information processing have many specific characteristics that require extensive and interdisciplinary investigation.

Along with the distinctive focus on the general properties of the human mind, cognitive psychology has contributed a particular method of investigation to cognitive science. Cognitive psychologists specialize in testing theories by making systematic, precise observations of human behavior, often under laboratory conditions. Several discussions in the chapter include detailed examples of laboratory experiments. In learning to think like a cognitive psychologist, you should do more than simply absorb the theory of cognitive architecture that is being presented. You should also consider and try to evaluate the experimental evidence discussed. When the evidence for a theory seems weak to cognitive scientists, they try to devise a further experiment that will either support the theory or weaken it in favor of some alternative. Cognitive psychology depends on the interplay of theory and systematic empirical evidence.

2.2 The Notion of Cognitive Architecture

A fundamental view in cognitive science is that an intelligent system is not completely homogeneous. It must consist of a number of functional subsystems, or modules, that cooperate to achieve intelligent information processing and behavior. This view is supported by results in cognitive psychology, some of which are presented in this chapter and the next, and by discoveries that animal brains contain numerous specialized regions, some of which will be described in chapter 7. It is also supported by attempts to build intelligent devices. Computer designers and AI researchers have found that they have to impose a great deal of structure on their machines and programs in order to achieve powerful information processing.

Our particular concern in this chapter is a kind of structure called *cognitive architecture*. Cognitive architecture refers to information-processing capacities and mechanisms of a system that are built in. The architecture of a system may or may not give it the potential to acquire information-processing capacities that are not specifically built in. For example, the architecture of the simplest electronic calculators fixes their mathematical capacities, whereas the architecture of programmable calculators allows them to acquire new capacities by being programmed. Although the new capacities are not functional until the programs are loaded into the machine's memory, the potential for them is inherent in its architecture. If we understand the architecture, we understand not only what the machine does now but the range of things it can potentially do and how it is able to do them. Given the responsiveness of human cognition to experience, the desire to understand the human cognitive architecture makes sense. To the extent that we understand it, we understand not just what some individual or group is capable of at a particular time and place but rather the human cognitive potential.

Since most adult cognitive abilities depend on extensive learning experiences, it is difficult to uncover the contribution of the cognitive architecture to these abilities and separate it from the contribution of specific experiences. The high degree of flexibility of human cognition requires that we think of much of the human cognitive architecture not as determining specific thoughts and behaviors but as an abstract set of mechanisms that potentiates a vast range of capabilities. We must get underneath people's behavior in particular circumstances to discover the basic information-processing capacities that allow adaptive responses to a wide range of situations.

Individual Differences in Cognitive Architecture

It should be noted that we make the simplifying assumption here that the human cognitive architecture is uniform across individuals. In fact, there is some architectural variation across individuals. The nervous system is subject to genetic variation, its early development can be altered by various kinds of nutritional or experiential deprivation, it can deteriorate with age, and it can be injured. In principle, variations can be qualitative or quantitative. Qualitatively different architectures have different information-processing components or structures. For example, a person who is totally deaf and a hearing person do not have the same cognitive architecture. The cognitive architecture of the hearing person includes a subsystem that provides acoustic input. That subsystem is functionally absent in the deaf person. In other respects the two people may have the same cognitive-architectural structures. In fact, enough of the deaf person's auditory architecture may be intact to allow considerable auditory cognition, even though the person cannot respond to sound. A case in point is Ludwig van

Beethoven, who composed his late works after going deaf. So-called color blindness is a more subtle example of a structural difference. A person with normal human color vision possesses three visual input channels for daylight vision, which have peak sensitivities in three different regions of the visible spectrum. A color-blind person might be missing one of these input channels and therefore be unable to make as many color discriminations as the person with three channels (Nathans 1989). An example of a quantitative difference in cognitive architecture is the deterioration of night vision with age. Owing to aging processes that are universal and currently irreversible, older people cannot see as well as younger people in the dark. The part of their visual architecture that is responsible for night vision operates qualitatively in the same fashion as that of younger people, but they are much less sensitive at low illumination levels and much slower to adapt to sudden changes in illumination.

We bypass variations in the architecture across individuals for two reasons. First, most human beings appear to have structurally very similar architectures. The variation is mostly quantitative. Once a structure is well understood, its quantitative variations tend to be easy to understand. Second, understanding what might be called the generic human architecture is a good way to begin to discover and try to understand the qualitative variations that do exist. Of course, there are cases where the observation of a structural variation can lead to new insights about cognitive architecture. Studies of so-called split-brain patients (introduced in chapter 7) are one such case.

The tendency of cognitive scientists to focus on the fundamental structure of the cognitive architecture tends to surprise students whose initial interest in human cognition arises from curiosity about cognitive differences among people. Different people have different abilities and different styles of thinking. It is natural to be curious about possible inborn differences that facilitate the development of unusual cognitive abilities, in the arts, sciences, practical affairs, and so on. In complex societies, which contain many demanding and competitive environments, such differences are also of practical importance. To cognitive scientists, however, the question of inborn human uniformity is deeper and more fascinating. A key contribution of cognitive science has been to bring out the interest and scientific importance of what we tend to take for granted: that almost all people have remarkably similar cognitive potential and even remarkably similar developed cognitive capacities. Nearly everyone is able to negotiate the visual world with great success, learn a natural language, and acquire a vast store of knowledge that can be deployed flexibly to meet the demands of an ever changing environment. Although we do not yet know everything we would like to about the limits and variability of human skill acquisition, it is also safe to say that with suitable practice nearly everyone could become competent at a wide range of cognitive and sensory-motor skills, from baseball, to cooking, to managing a preschool classroom, to solving physics problems, to leading a nomadic life in the desert.

This perspective on human cognition also leads toward a perspective on human differences that does not overemphasize genetic variability. To a very large extent the differences in people's cognitive abilities are the result of differing goals and experiences. It is no doubt true that if environment and experience could be held constant, people's achievements would vary, in part because of inborn cognitive differences, but these variations can be seen as the wrinkles in the rich and never completed tapestry of human accomplishment that is woven from the interaction between the basic cognitive architecture and the variety of human environments.

In a sense, the root problem of cognitive science is to understand how the cognitive architecture makes mundane human competence possible. An analogy with the history of chemistry is useful here. The alchemists did not make much progress when they focused their attention on exotic problems such as transmuting lead into gold. Chemistry advanced when thinkers began to focus on the underlying architecture of all matter. From this point of view mundane substances such as table salt and water are as interesting as gold, and understanding ordinary chemical reactions, such as the burning of a candle, holds the key to progress. Following this lead, we will concentrate on the ordinary in cognition.

2.3 A Global View of The Cognitive Architecture

In order to focus our discussion in this chapter, our first task is to break the cognitive architecture into subsystems. Everyday intuition, evidence about the brain, and evidence from laboratory studies in cognitive science are in good agreement about at least some of the major components of cognition. Human beings can receive information through their senses, think, and take physical action through voluntary muscular movement. As it is depicted in figure 2.1, then, we can initially conceive of the human information processor as consisting of a central thinking system that receives information about the world from sensory systems and issues movement commands to the motor system. In addition, language plays a large and biologically unique role in human cognition. It is shown as a specialized subsystem of the central cognitive processor in figure 2.1. Central processes are the subject of this chapter. The linguistic component of the cognitive architecture is treated in chapters 6, 9, and 11. In chapter 12 we take up vision as an example sensory system. Chapter 7 concerns the neural implementation of the architecture.

The decomposition of the human information-processing system shown in figure 2.1 should be taken as illustrative and provisional at this point. The senses (vision, audition, and so on) could be represented as separate subsystems, raising questions about whether they intercommunicate directly or only via the central system. Evidence might lead us to split the central system up into a number of components with distinctive channels of interaction. On the other hand, it must be recognized that the

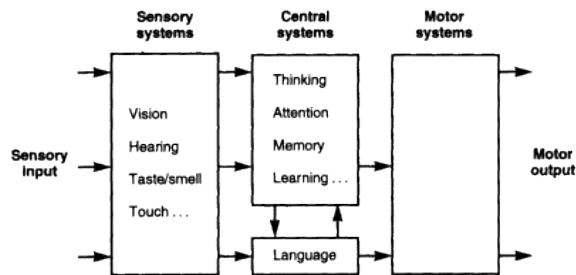


Figure 2.1
A global view of the cognitive architecture.
(Used with the permission of Neil Stillings.)

boundaries between modules and the channels of communication may not be so clear-cut. The separation between central processes and the sensory and motor systems, for example, is not as sharp as the figure indicates. After sensory information is initially registered, it is operated on and transformed successively, making it difficult to identify where sensory or perceptual processing stops and central processing begins, or what portion of the transformation from sensory input to motor output is to be designated central. Similarly, the boundaries between the linguistic system and sensory, motor, and other central systems is not completely clear. Some of the fuzziness will probably be resolved by future research, but it may also be necessary to recognize that any simple partitioning of the cognitive architecture is partly an intellectual convenience that obscures some of the complexities of human information processing.

Input and Output Systems

Each major component of the cognitive architecture has particular functions. The first job of a sensory system is the *transduction* of some form of physical energy impinging on the body into a form that usable by cognitive processes. At the physical level of analysis, the incoming physical energy is transduced into electrochemical activity in specialized nerve cells called *receptors*. At the formal level of analysis this electrochemical activity can be looked at as a complex signal or code that represents information contained in the incoming physical energy. Typically the initial encoding must be transformed in various ways to produce a clear representation of the most useful information. Sensory systems have elaborate architectures to accomplish the initial transduction and further processing of sensory input. In the visual system, for example, the retinas at the backs of the two eyes contain millions of *photoreceptor* cells. The initial encoding delivered by the photoreceptors is processed extensively by several other layers of retinal cells and by many specialized regions in the brain.

The specialization of sensory systems is one source of evidence for the *modularity* thesis concerning the cognitive architecture (Fodor 1983). A prominent feature of this thesis is that sensory systems are *informationally encapsulated*. That is, a sensory system transforms a class of environmental inputs into outputs that are useful to central processes, but the computation of its output is not affected by central processes. Input systems are thought to have evolved to rapidly deliver reliable information to central processes, and their structure and operation is argued to be genetically determined and not substantially modifiable. The modularity thesis amounts to a claim that some of the boundaries among the major components of the cognitive architecture in figure 2.1 can be drawn in a principled way using a combination of computational and biological criteria.

Fodor (1983) argues that the initial processing of spoken linguistic input is also modular. Given the intimate connection that we sense between language and thought, it seems strange at first to think that the processing of linguistic input could be informationally encapsulated. Knowledge and expectations that are available centrally might well affect the perception of language. There is considerable debate about this issue, but regardless of the outcome of that debate later chapters will make clear that there is much that is biologically and computationally special about language.

Similar points could be made about the motor system. It also involves transduction. In order for actual movement to occur, neural codes must cause muscular contraction. This transformation is accomplished by neurochemical events at a specialized type of connection between motor nerve cells and muscle cells called a *neuromuscular junction*. As with sensory systems, several parts of the brain are specialized for motor control.

Central Systems

Our primary job in this chapter and the next is to characterize the architecture of the central information processes. We will assume that the sensory and linguistic systems can deliver meaningful input to the central processes. For example, we assume that the visual system can reliably deliver the spatial locations, shapes, colors, and identities of common objects in the visual field. And we assume that the linguistic system can reliably identify the words and the grammatical structure of incoming sentences. We assume that the motor system responds reliably to motor command sequences issued by central processes by moving the body in the appropriate ways. The execution of motor commands includes moving the muscles of the vocal tract to produce speech. So, given that the central system has appropriate inputs and outputs, we ask how it is organized to accomplish the aspects of cognition that we associate distinctively with it.

Everyday terms for some of the capabilities of central cognition are listed in figure 2.1. First and foremost, we can think. That is, we can reason from known facts to novel conclusions, we can solve problems, we can plan our activities and evaluate alternative plans, and so on. Furthermore, our thinking capacity is remarkably flexible. It is often not controlled by immediate sensory input or the need for immediate action. There is no obvious limit to our thinking, in the sense that we draw new conclusions, make new plans, and solve new problems in remarkably diverse domains every day. Thinking is usually characterized by a focus of attention. The positive side of attention is that it seems to maintain the goal-directedness of thought. The negative side is that our attentional capacity seems to be limited and is sometimes overwhelmed. In addition to drawing on current perceptual input, thought also takes inputs from memory. Human memory, though far from infallible, seems relatively unlimited both in capacity and in ability to yield up information that is relevant to current thought processes. Finally, there is our capacity to learn. We acquire the knowledge that resides in our memories, and we are able to acquire many cognitive and sensory-motor skills during our lifetimes.

It is apparent that central processes must have a rather complex architecture of their own. In the 1980s a view of the architecture of cognition came to a kind of maturity after a long period of development that began in earnest in the 1950s. We will refer to this view, perhaps a bit too grandly, as the *classical* view (Fodor and Pylyshyn 1988). In truth, it would be more accurate to call it a family of views, since there have been a number of somewhat different architectures proposed that are all recognizably classical (Anderson 1983; Newell 1990; Simon 1990 are examples). Much of this chapter will be devoted to developing a generic version of the classical view. The chapter will also introduce a significant critique of the classical view and a family of new proposals about cognitive architecture that have arisen over the last ten years. These views go under a number of names, including *connectionism*, *parallel distributed processing*, and *artificial neural networks*.

Physical Symbol Systems

A key hypothesis in the classical view of the cognitive architecture is that the central processes of the human mind constitute, among other things, a general-purpose computational system, or *physical symbol system* (Newell 1980; Newell, Rosenbloom, and Laird 1989). The hypothesis includes the framework developed in chapter 1. That is, it is assumed that cognition can be analyzed as formal symbol manipulation processes, which can be given a semantic interpretation that accounts for their meaningfulness.

and adaptiveness to the environment. The symbol system is physical because the formal processes are implemented in a particular physical medium, the brain. The additional content of the hypothesis has to do with the range of symbolic processes that are made possible by the cognitive architecture. The idea is that the architecture contains built-in structures and operations that make central cognitive processes programmable. The architecture contains a set of built-in information-processing facilities that endow it with the potential to acquire an unlimited range of further information processes.

Universal Machines Research in computer science and allied mathematical fields has led to an understanding of properties that a flexible programmable system seems to require. The system must have a memory in which symbols can be stored and from which they can be retrieved. Maximum flexibility requires that the size of the memory be arbitrarily extendible, since there is no limit to the amount of stored information that a task might require. Extendibility is made possible in computers through the provision for adding ever more or larger mass storage units, such as disk and tape drives. Human memory is extendible through external storage media such as handwritten notes, books, tapes, and computerized records. The storage capacity of the human brain is also extremely large, so that its limits are not exceeded for many tasks.

The processor must be able to store symbols in memory and retrieve them. Moreover, symbolic processing requires that what is stored in memory can depend on what has recently been retrieved. This allows the processor to systematically produce new symbolic structures in response to those already in memory or to transform structures already in memory. For example, a calculator might be able to retrieve a pair of symbols from memory and produce a third symbol according to an algorithm that represents addition.

A truly flexible system is able to acquire new information-processing procedures. In conventional programmable systems some of what is retrieved from memory functions as instructions to perform certain operations rather than merely as data to be fed into operations. For example, a programmable calculator might be able to retrieve a triple of symbols from memory and treat the first symbol as an instruction to perform some arithmetic operation on the second two symbols. In computer science, the process of treating information stored in memory as instructions is often referred to as *interpretation*, since the system is, in a sense, interpreting the information as instructions. This sense of interpretation is obviously different from the kind of semantic interpretation discussed in chapter 1. The interpretation function inside the machine establishes a relation between formal symbols and formal operations, whereas semantic interpretation establishes a relation between formal symbols and operations and some significant domain external to the system.

In the 1930s Turing (1937), an early computation theorist, gave one of the first general characterizations of the class of symbolic transformations that can be computed by formal processes. He defined a class of devices, now called *Turing machines* (described in more detail in chapter 8), each of which computed some input-output function. He called the entire class of functions computable by Turing machines the *computable functions*. (He also proved the existence of functions that are not computable.) The argument (known as *Church's thesis*, *Turing's thesis*, or the *Church-Turing thesis*) that the computable functions are a natural class has stood the test of time in the sense that all other proposed universal characterizations of computation have been

shown to define the same class of functions. A device capable of computing any computable function is known as a *universal machine*. Turing defined a universal Turing machine, for example, that could mimic the computations of any other Turing machine. The mimicry of another machine was accomplished by storing a description of it on the universal machine's memory tape and interpreting the description. The description of the other machine is essentially a program. A consequence of Turing's results is that if a system has a memory and the ability to interpret the contents of its memory as instructions to perform operations, the repertoire of built-in operations needed for universality is extremely simple.

The universal Turing machine, although a convenient mathematical fiction, is not suitable for practical computation and hence has never been built. Its extremely inefficient memory retrieval scheme requires scanning serially through arbitrarily long sections of its memory tape to locate needed information or to find the correct cells to write in. Its primitive operations are so simple that considerable programming is required to build up operations such as addition or comparing two strings of symbols for equality. Because of the inefficiency of memory storage and retrieval, these relatively elementary operations would be extremely inefficient, making complex programs that depended on them fatally slow.

From the late 1940s through the present the *von Neumann architecture*, named after the mathematician John von Neumann, has been employed in most practical general-purpose computers. The hallmark of a von Neumann machine is a large *random-access memory*. Each cell in the memory has a unique numerical address, which can be used to access or replace the contents of that cell in a single step. In addition to its ability to address memory locations directly, a von Neumann machine also has a central processing unit (the *CPU*) that possesses a special working memory (*register memory*) for holding data that are being operated on and a set of built-in operations that is rich in comparison with the Turing machine. The exact design of the central processor varies considerably, but typically includes operations such as adding two binary integers, or branching to another part of a program if the binary integer in some register is equal to zero (so-called *conditional branching*). The CPU can interpret information retrieved from memory either as instructions to perform particular operations or as data to apply the current operation to. Thus, a portion of memory can contain a sequence of instructions, called a *program*, and another portion of memory can contain the data to be operated on by the program. The CPU repeatedly goes through a *fetch-execute* cycle, in which the next operation and its arguments are retrieved from memory and the operation is applied to the arguments. Although it computes the same class of functions as a universal Turing machine, a von Neumann machine runs efficiently because of its random-access memory and because its architecture can be implemented in electronic circuitry that makes its basic operations extremely fast (currently from millions to tens of millions of instructions per second in desktop computers).

Although it is immensely practical, a von Neumann machine, like the universal Turing machine, might be called an immaculate universal device. It has the potential to do anything, but as it comes out of its shipping carton it actually can do nothing—until a program is loaded into its memory. The noncommittal purity of von Neumann machines is a major source of their economic success. The machines are engineered precisely to be bendable to the demands of any information-processing task and therefore have the widest possible market. The manufacturer wastes no money building in functions that many users would not want. The flexibility of a von Neumann

machine, however, can only be realized by an intelligent programmer. It is a tool, perfectly adapted to an environment of programmers but lacking any innate autonomy. The clean separation between hardware and software, where the responsibility of the hardware is to provide efficient and cost-effective computational potential via programmability, has been a spectacularly successful technological strategy.

This strategy obviously would not suffice for biological success. Organisms must be innately endowed with specific computational capacities that are dedicated to tasks necessary for their survival. In addition, more intelligent organisms, including humans, can acquire new information-processing abilities by interacting with their environments. They can, in a sense, be programmed by a sequence of goal-oriented experiences in their environments. Neither dedicated computational capacity nor learning from experience is typical of physical symbol systems as they have been developed technologically. The fact that physical symbol systems need have no built-in active survival mechanisms demonstrates that symbol-systemhood is not a sufficient condition for intelligence. One claim of the classical view, however, is that being a physical symbol system is a necessary condition for intelligence at the human level. At the heart of this claim is the notion that the extraordinary behavioral flexibility demonstrated by human beings can only be achieved by a physical symbol system.

It seems clear that nervous systems that function as full symbol systems are not widespread in the biological world. Solutions to particular computational problems abound in evolution, for example, the frog's visual prey-detection system, the bat's sonar localization system, the electric signaling system of the electric fish, the housefly's flight-control system, or the song-acquisition mechanisms of various species of songbirds, but only the human nervous system seems to be a full symbol system. Newell (1990) refers to the shift from dedicated hardware for specific computational functions to a general-purpose physical symbol system as the *great move*. It is not clear just why humans exhibit so much more cognitive flexibility than other intelligent animals, such as the other primates. Their cognitive architectures are probably both qualitatively and quantitatively different from ours, but the details of the differences are far from clear. Just how human cognitive flexibility emerged from evolution's generally more focused ways is also not clear. Given the differences between biological information processing and computing technology as it has developed so far, we cannot look to the details of the von Neumann architecture for insight into the architecture of the human central processor. There is no evidence, for example, that human memory consists of discrete, numerically addressable locations or that the primitive, built-in operations of human thought are the same as those of a von Neumann machine. It is a matter of empirical research to discover how a physical symbol system is implemented in the human mind and brain. The physical symbol system hypothesis makes no prior commitment about the details.

Computation in Symbol Systems The hypothesis that the mind is (among other things) a physical symbol system is highly abstract. There has been considerable discussion about just how it constrains theories of cognition. A basic requirement is that the cognitive system be able to compose new functions or transformations over its representations. Newell (1990) has pointed out that there are many ways that this capacity might be implemented. The von Neumann solution of using the same medium for representing both data and instructions is elegant but only one option. Another option might be to implement new functions directly in reconfigurable hardware, essentially

building special-purpose subsystems to compute new functions as needed. The notion of interpretation would shift quite radically in such a system. Instructions, in some sense, would be needed to build the special-purpose subsystem, but once built the subsystem would run directly, without the interpretation of instructions retrieved from a memory. Some points about the classical architecture made below can be seen as a claim that human cognition includes both von Neumann—style interpretation and a mechanism for constructing new functions that operate without interpretation.

The memory retrieval processes in a symbol system allow what Newell (1990) calls *distal access* to symbols. Formally, what this means is that the processing of one symbol or symbolic structure can reliably evoke the processing of other symbolic structures that are needed in the current computation. A common example is when a single symbol serves as an index or pointer to a larger symbol structure with which it is associated and which can be brought into the computation. In standard computer programming languages when the name of a procedure or subroutine is encountered during a computation, it triggers the application of the entire subroutine. In human language understanding the occurrence of any well-understood content word potentially triggers the processing of extensive symbol structures that represent the meaning of the word. Newell argues that the need for distal access reflects a constraint on the physics of computational devices. The physical implementation of any information-processing system imposes a limit on the amount of information that can be involved in a computation at any moment. The limit can be overcome by passively storing large amounts of symbolic structure in regions of the system that are physically distinct from the regions that are actively supporting the computation and by arranging a mechanism that allows the passive structures to be brought into the computation when needed.

The ability of a symbol system to maintain internal representations of the world, to access them, and to transform them with processes that are not immediately driven by sensory input or tied directly to motor output is crucial to its flexibility. The physical symbol system hypothesis is motivated in part by the human capacities for planning, problem solving, and reasoning that are not driven by the immediate sensory input and that make novel use of large stores of knowledge.

Symbol systems' ability to process syntactically structured symbols is also widely considered to be crucial to their flexibility and power. Although facilities for handling complex syntactic structure are not transparently built into the simplest universal architectures, such as the Turing machine, or even into standard von Neumann computers, programs for building, recognizing, and transforming symbolic structure can be built up from the primitive operations in such architectures. Universal computational capacity seems to require such facilities. Fodor and Pylyshyn (1988) argue that structured symbols and *structure-sensitive* computation are defining characteristics of human cognitive processes. The human cognitive architecture, they argue, has an intrinsic capacity to build up complex symbols out of simpler ones, to identify the parts of complex symbols and their arrangement, and to carry out computations that depend on symbolic structure. These capacities are very apparent in natural languages. English noun phrases, such as *the old man*, are a convenient example. *The old man* is itself a complex symbol, as can be seen from related structures, such as *the man* or *the young man*. It can also serve as a unitary part of larger structures such as *The old man kicked the uncooperative donkey*, as can be seen from related structures such as *Irving kicked the*

uncooperative donkey, *The uncooperative donkey kicked the old man*, or *The donkey who the old man kicked was uncooperative*. The structure-sensitivity of natural language processing is brought out in these examples by the fact that who did the kicking and who was kicked cannot be determined without identifying the grammatical structure of the sentences. *Kick*, *man*, and *donkey* are present in three of the sentences, but the kicker and the kickee are not the same in all three. (The structure of natural languages is taken up in detail in chapter 6.)

Structure-based computation makes possible the *productivity* and *systematicity* of symbolic computation. Complex symbols are built up and processed according to sets of rules or constraints. The rules can be applied repeatedly to produce or analyze ever more complex symbolic structures. To return to the example in chapter 1, a system that contains rules for dealing with arithmetic expressions can deal not only with expressions, such as $3 + 2$, but also with arbitrarily complicated expressions, such as $(3 \cdot (5 + 7)) + (6 \cdot 9)$, in which the order of application of the arithmetic operations depends on the structure of the expression. Similarly, the rules in natural languages for forming noun phrases and combining them with other types of phrases allow for the generation and processing of an unlimited stock of novel sentences, which can be arbitrarily complex. Claims about the productivity of a system are somewhat abstract, because the ability of a system to apply a set of rules or constraints is limited by its finite computational resources. For example, a calculator with 64,000 memory cells and no auxiliary storage might not be able to process an arithmetic expression containing 100,000 symbols. Similarly, the grammatical rules for English allow for sentences that would be too long to produce or process in a lifetime, and even some relatively short sentences, such as *The old man who the uncooperative donkey that the circus trainer sold kicked died*, raise interesting questions about limits on the capacity of natural language processing. In spite of these limits, it can be argued that in domains such as natural language or musical composition within a style human cognition is spectacularly productive.

Structure-based computation is also systematic. Information processes operate on representations in terms of their structure, guaranteeing that representations with similar structures will be processed in the same way. A system that employs the rules of arithmetic will not deal correctly with $3 \cdot (5 + 7)$ but balk at $7(5 + 3)$ or $(3 + 2) \cdot (5 + 7)$, for example. Similarly, a person who knows English vocabulary and grammar easily understands novel sentences, such as *John kissed the skunk*, and, having understood a sentence of a particular form, will rarely fail to understand other utterances with the same syntactic form, such as *The skunk kissed John*. Even nonsense utterances that contain cues to grammatical structure, such as *The wug scrogged the blap*, receive considerable processing relative to syntactically unstructured strings such as *wug scrog blap*.

As we pointed out in chapter 1, the meanings of structured representations can be specified in terms of their syntactic structure. The semantic mapping from *John kissed the skunk*, for example, requires that the person specified by the grammatical subject of the sentence must have a certain relationship to a skunk in the world. The required relationship is reversed in *The skunk kissed John*. This type of *compositional* semantics is hypothesized to be characteristic of human language and thought.

In recent years the physical symbol system hypothesis has become controversial in cognitive science. A particular subject of attack has been the detailed claim that

rule-governed processing of syntactically structured representations is central to human cognition. In the rest of this chapter we will first develop the classical view further and then sketch the alternative views that have arisen under the name of *connectionism* or *parallel distributed processing*.

2.4 Propositional Representation

The Language of Thought

Preliminary Arguments for Propositional Representation We begin our further exploration of the classical view of the cognitive architecture with one of the fundamental questions that has occupied its proponents. What "language" is our factual, or *declarative*, knowledge represented in? In writing a computer program, the programmer specifies a representational format for the data structures. For example, in an employee information system the facts about each employee might be stored as a record, containing slots for name, address, phone number, title, monthly salary, number of dependents, seniority, and so on. The format for a record is an abstract minilanguage for specifying the information about the employee. What, then, can we say about the much more general ability of the human mind to represent an unlimited variety of facts in an unlimited variety of subject areas?

One possibility is that facts are represented solely in the thinker's native *natural language*, for example, English or Chinese. This possibility is attractive. For one thing, we clearly do possess linguistic representations, since we produce and understand them all the time. And natural languages are clearly powerful enough to represent an unlimited variety of facts, unlike artificially restricted personnel record languages and the like. Introspectively, it often seems that we think in our native languages as well. We will not embrace this theory, however. Instead, we will assume that facts are also and primarily represented in a still more powerful internal *propositional* representation that is largely unconscious and that is translated into natural language when we talk.

Although it is difficult to fully justify this fundamental assumption in advance, there are some introspective arguments and experimental evidence that justify adopting it at least provisionally. Much of the material in the rest of the chapter can also be seen as further support for it, although that point will not be emphasized (see also Fodor 1975).

The theory of an underlying propositional representation corresponds to the frequent introspective sense of a distinction between having an idea and putting it into words. The distinction emerges behaviorally in the tip-of-the-tongue phenomenon, where we have a concept clearly in mind but experience a delay in retrieving the appropriate word for it. Furthermore, even in cases where we easily generate the English name for a concept (*dog*, *chair*, and so on), we often are not very good at giving its English definition. When asked to define even the most familiar concepts, we often find ourselves laboring. This phenomenon suggests that the definitions are expressed in an internal representation that is associated with an English word. We retrieve the word effortlessly, but a full translation of the definition into English is not immediately available and requires considerable mental effort. A final reason for at least initially adopting the internal representation theory is that it explains how young children can begin acquiring concepts in spite of the fact that they are not born knowing a natural language. Whatever we might think of adults' abilities to define concepts in their

native languages, children seem easily able to learn concepts that they will be completely unable to explain until they have been talking for quite some time.

Experimental Evidence for Propositional Representation A now classic experiment by Sachs (1967) is an example of the kind of laboratory evidence that supports the theory of an underlying propositional representation. In the experiment subjects listened to paragraphs on various topics, such as Galileo's work on the telescope. The passages were interrupted, and subjects were given a *recognition memory* test for a sentence that had occurred 0, 40, or 80 syllables earlier in the passage. In the test the subject was presented with a sentence and asked to judge whether it was the same as a sentence in the passage or changed. For the Galileo passage the subject could be presented with any one of the following four sentences. The actual sentence was sentence (1).

- (1) He sent a letter about it to Galileo, the great Italian scientist.
- (2) He sent Galileo, the great Italian scientist, a letter about it.
- (3) A letter about it was sent to Galileo, the great Italian scientist.
- (4) Galileo, the great Italian scientist, sent him a letter about it.

Notice that sentences (2) and (3) have the same meaning as sentence (1), although they have different grammatical forms. Sentence (4) has a different meaning. At a delay of 80 syllables (20—25 seconds) subjects rarely made the error of thinking that sentence (4) had occurred in the passage. However, they frequently and mistakenly thought they remembered sentence (2) or sentence (3). Thus, under the conditions of Sachs's experiment subjects were able to remember the meaning of a sentence without remembering its linguistic form. A reasonable explanation of the finding is that the meaning was represented in an underlying nonlinguistic form that could be translated into a variety of linguistic paraphrases, which subjects confused in the recognition test. The finding actually accords well with our frequent feeling that we can remember the gist of what someone has said but not the exact words. The experiment demonstrates quantitatively that the representation of exact wording can be lost remarkably quickly under quite normal listening conditions. We will return to the analysis of memory experiments in chapter 3, but it should be pointed out here that linguistic representation obviously can be remembered. One does not sing paraphrases of "The Star-Spangled Banner," for example.

Propositional Form If we adopt a theory of underlying propositional representation, we take on the imposing problem of figuring out in detail what the formal structures of propositions are. We will propose only the most rudimentary structures here, because the experimental methods of the cognitive psychologist are not the best tools to apply to the detailed analysis of the problem, initially. As a starting point, we assume that propositions are the simplest complete units of thought. They are complete in the sense that they can have a truth-value; that is, they can be true or false. Anything simpler seems incomplete in this sense. For example, *Mary* is not a complete proposition, nor is *likes*, but *Mary likes John* is because it seems to express a complete thought that can be either true or false.

Propositions typically capture relations, such as *liking*, that hold between arguments, such as *John* and *Mary*. *Like* is a two-place relation, or predicate, that requires two arguments to form a proposition. The arguments are not interchangeable: *Mary likes*

John and *John likes Mary* are distinct propositions. We can distinguish among the arguments of a predicate by noting the roles that they play in the proposition. *Mary* is the subject and *John* the object of *Mary likes John*. There are predicates with one place, such as *red*, predicates with three places, such as *give*, and so on. Two or more propositions that share arguments can often be expressed in a single sentence: for example, "Mary likes the teacher who gave her an apple," which is built out of a *liking* proposition and a *giving* proposition.

Propositions and their parts were placed in italics above to distinguish them from linguistic representation. Theorists usually also adopt special formal notations for propositions that allow them to be processed by formal procedures and that allow a formal theory of meaning to be specified. We will informally introduce a simple version of *propositional network* notation, which involves the theoretical commitments that are most important in cognitive psychology. Figure 2.2 shows propositional network notations for the propositions just discussed.

In this notation each proposition is a unique structural unit built around a *propositional node*, which is symbolized by a circle with a number inside it. The value of the number has no meaning other than to distinguish the proposition from other propositions. The propositional node is connected to its *constituent* nodes by *links*, or *pointers*, which are symbolized by arrows. Each link has a label that specifies the role that its target plays in the proposition. Every propositional node must have a link to a relation node, which is symbolized by a word in capital letters. The propositional node must also have the right number of links to nodes for the arguments of the relation. Argument nodes are symbolized by small letters when they represent particular individual entities in the world, such as the particular person, Mary. In some cases a previously anonymous entity must be assigned a name, such as *a* for the apple that John gave to Mary. Argument nodes are symbolized by capital letters when they represent general concepts for objects, or categories that objects belong to. *APPLE* and *TEACHER* are examples in the figure. The relation *IS-A* is used for propositions, such as *John is a teacher*, that assert that a particular entity falls in a certain category.

In the four propositions in part (a) of figure 2.2 some of the argument nodes occur more than once. For example, the node *john* occurs in three of the propositions. Part (b) of the figure displays the interconnections among the propositions more clearly by displaying each node once. Here, we see immediately that *john* is a structurally central node in the total network, because it is immediately connected to three of the four propositions. Note that the diagram of a proposition can be rearranged at will as long as the links still connect the same nodes. The differences between part (a) and part (b) of the figure demonstrate this point. Since our primary interest will be in the interconnections among propositions, rather than in their internal structure, the labels on the pointers have been suppressed in part (b). Readers interested in pursuing questions about the use of networks to represent the internal details of propositions should consult Johnson-Laird, Herrmann, and Chaffin (1984). We give a more formal analysis of propositions in chapter 10.

Experimental Evidence for Simple Propositional Units There is good experimental evidence that propositions are important functional units in cognitive processing. One preliminary example is an experiment by Kintsch and Glass (1974). They compared memory for sentences that expressed a single proposition with memory for sentences with the same number of content words that expressed more than one proposition. For

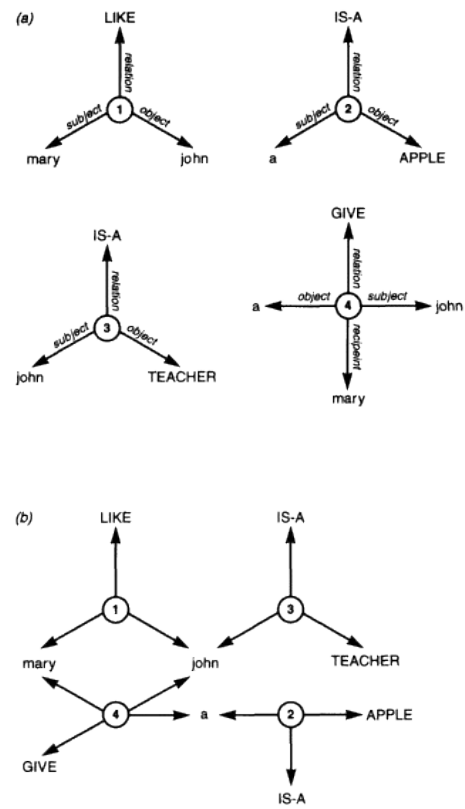


Figure 2.2

Network notation for propositions. (a) Notation for four propositions with all links labeled. (b) The four propositions combined into a single network. The link labels are suppressed.

example, sentence (5) is constructed from one proposition, involving the three-place relation *BUILD*, whereas sentence (6) is constructed from three propositions, involving the three one-place relations *CROWDED*, *SQUIRM*, and *UNCOMFORTABLE*:

(5) The settler built the cabin by hand.

(6) The crowded passengers squirmed uncomfortably.

Subjects tended to recall the first kind of sentence better than the second. This indicates that it had more unity in an underlying representation that determined the subjects' memory performance.

Propositional Networks as Associative Networks

A number of further assumptions must be introduced to make propositional networks a full-fledged theory, or *model*, for long-term declarative memory. *Long-term memory* refers to cases where a piece of information remains out of the immediate sphere of consciousness for some reasonable length of time (from minutes to years) and then is *retrieved*, or brought back into conscious attention, when it is relevant to some ongoing thought process. The basic assumption of the theory is that at each moment in time each node in a network is at some level of *activation* and that activation spreads among nodes along the links. If the level of activation reaches a high enough value in some portion of the network, that portion of the network is accessible to conscious attention. The links are therefore *associative* connections that determine the tendency of one item of information to lead to another in thought.

Our experience of the stream of consciousness in everyday life supports this notion of *spreading activation*. If we relax our attention to current goals, one idea seems to lead to another with no effort, and the connections among the ideas, although meaningful, are much looser than would be dictated by a rationally controlled and goal-directed process. The flow of ideas seems to reflect a rich network of associations among facts and experiences stored in long-term memory. Virginia Woolf captured this mode of experience in her novels, and psychoanalysts use it to try to uncover associations that are thought to be suppressed by the ego's mechanisms of defense. In more goal-directed thought we are able to pick and choose from the activated material the information that is relevant to our goal, and we are able to increase and maintain the activation of this material by attending to it.

Although the theory of association of ideas goes back to Aristotle and was developed vigorously by nineteenth-century philosophers and psychologists, the information-processing approach has made possible the development of versions of the theory that make novel qualitative and quantitative predictions. Most versions of the theory make several assumptions in addition to the assumption that activation spreads through a network along the links. First, it is assumed that activation spreads simultaneously, in *parallel*, to all of the links leading from a node, and that it can spread in either direction along the links. Second, it is assumed that the activation at a node fades rapidly with time. Third, it is sometimes assumed that the total activation at a node is divided among the links that lead from it. Thus, as activation spreads through the network, it is weakened by successive divisions as it reaches successive nodes. The second and third assumptions prevent a single thought from rapidly activating the entire long-term memory permanently, throwing it into a state of permanent confusion. That is, the usefulness of spreading activation depends on some relevant nodes being in a higher state of activation than the other nodes in a network. Fourth, it is

often assumed that nodes and links can have different activation capacities. For example, links might be assumed to differ in the proportion of the available activation they receive, where the higher-capacity links are the ones that have been more frequently activated in the past. Quantitative predictions become possible when concrete assumptions are made about how rapidly activation spreads, how rapidly it decays, and so on (Anderson 1983; Ratcliff and McKoon 1981).

Propositional networks are examples of richly interconnected information structures that can be processed in parallel. The spreading activation mechanism is parallel because the activation process affects many parts of a network simultaneously. The theory is one proposal for a very simple parallel computational mechanism that can efficiently make parts of a network available for further processing. It agrees with our intuitions from the stream of consciousness and also with overwhelming evidence that the nervous system is designed on the physical level for highly parallel information processing. The connectionist approach to the cognitive architecture, which will be discussed later in this chapter, arose in part from a desire to explore the properties of parallel computation in networks more generally.

Experimental Evidence for Propositional Networks The kind of evidence that has been obtained in the laboratory for propositional network structure is well illustrated by the work of McKoon and Ratcliff (1980). They exploited a prediction of spreading activation theory that is not obvious at first sight. Activation spreading from a source proposition raises the level of activation of other propositions. This should *prime* them for participation in later information processes. Consider the network in figure 2.2. Thinking about Mary, even just reading the word "Mary," should spread some activation to the *john* node, which will make it easier for this node to reach a level of activation sufficient for conscious processing. For example, the prior activation of the *mary* node might make it easier to read the word "John" or to answer the question "Is John a teacher?"

On each study-test sequence of McKoon and Ratcliff's experiment the subjects read two short paragraphs with simple propositional structures, which were presented on a computer screen. They were then given an immediate recognition memory test. Thirty-six test words appeared on the screen, one at a time. For each word the subjects responded as quickly as possible by pressing a "Yes" key if they thought the word had appeared in one of the stories or a "No" key if they thought it had not. On *primed* trials the test word was preceded by a test word from the same story. On *unprimed control* trials the test word was preceded by a test word from the other story. There were also "No" trials in which the test word had appeared in neither of the stories. Response times on primed trials were faster than response times on unprimed trials. The network theory predicts this finding, because there was always some associative pathway between any two words in a story. Associative pathways between two words from different stories would have to have been established prior to the experiment and would not be consistent enough across subjects to show up in the data.

McKoon and Ratcliff also constructed their stories to allow quite detailed predictions of the relative sizes of priming effects within stories. Here is one of their paragraphs:

The businessman gestured to a waiter. The waiter brought coffee. The coffee stained the napkins. The napkins protected the tablecloth. The businessman flourished documents. The documents explained a contract. The contract satisfied the client.

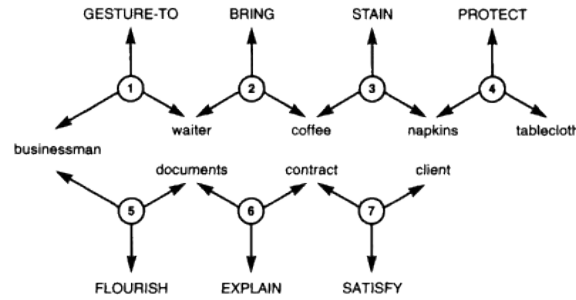


Figure 2.3
Propositional network for a paragraph.
Prime-test pairs were waiter-documents, coffee-contract,
and napkins-client. (Adapted from McKoon and Ratcliff 1980.)

This is a boring story with an exciting propositional structure, which is diagrammed in figure 2.3. (Note that for simplicity we have used some shorthand, for example, *waiter* rather than an *IS-A* proposition linking the concept WAITER with a name *w*.) In the story the following pairs of nouns are all exactly the same number of sentences apart: *waiter-documents*, *coffee-contract*, and *napkins-client*. The second word of each pair was primed by the first during the experiment. If the subject remembered the story word for word, the priming effect would be the same for all pairs. In the network structure, however, the distances between the pairs are quite different. There are four links and three nodes between *waiter* and *documents*. There are eight links and seven nodes between *coffee* and *contract*, and there are twelve links and eleven nodes between *napkins* and *client*.

The priming effect of *waiter* on *documents* should be quite strong, because activation is split up among pathways only a few times. It should be successively weaker for the other two pairs, because of the increasing dilution of the spreading activation along the longer pathways. The results were as predicted. The response time to *documents* was 736 milliseconds unprimed and 665 milliseconds when primed with *waiter*, giving a 71-millisecond priming effect. The priming effects for *contract* and *client* were 47 milliseconds and 30 milliseconds, respectively. (Actually, the response times are averages over forty-eight different paragraphs with the same propositional structure; the use of different paragraphs ensures that the results are due to the structure and not to peculiarities of the content of a particular paragraph.) In a second experiment McKoon and Ratcliff showed that similar priming effects could be obtained with more complex and realistic prose passages.

2.5 Schematic Representation

Conceptual Schemas

The theory that declarative knowledge is a network of propositions is incomplete as it has been developed so far in this chapter. The most serious gap is that although a representation for specific facts has been developed, no account has been given of the

general knowledge that allows the system to formulate or understand such facts in the first place. For example, in figure 2.2 we represented the fact that a certain object was an apple by storing an *IS-A* relation between *APPLE* and *a*. But this proposition is useful only because the mind also possesses general information about apples; that is, it has a general *concept* of apple. People know that an apple is an edible fruit that grows on a tree, has a roundish shape, is often but not necessarily red when ripe, and so on. Part of a theory of the architecture of the mind must be an account of such general knowledge.

Our initial hypothesis about the general knowledge for basic concepts is that it is represented in the propositional network. The network can easily contain general propositions, such as one asserting *ROUND* of *APPLE*, as well as propositions about particular objects, such as our example linking *APPLE* and *a*. The cluster of general propositions attached to a concept node forms a conceptual *schema* for the concept. We will use the term *schema* (plural is either *schemas* or *schemata*) for any cognitive structure that specifies the general properties of a type of object or event and leaves out any specification of details that are irrelevant to the type. A schema is an *abstraction* that allows particular objects or events to be assigned to general *categories*. General knowledge of the category can then be applied to the particular case.

The conceptual schema for apples specifies general information about fruithood, shape, color, and so on, but it leaves out many characteristics of individual apples, such as the precise details of their coloring, whether they fell or were picked, the supermarket where they were purchased, and so on. The schema *abstracts away* from the details in order to allow categorization and further thought and action based on the categorization. Some form of schematization is essential to intelligent information processing. Since every concrete object has an infinite number of characteristics, if we tried to deal with each object in all of its individuality, we would be permanently paralyzed and bewildered. Categorization also obviously has pitfalls, as well. Apples picked from the tree are less likely to be bruised, and those sold by certain supermarkets are more likely to be fresh. Allowing one's attention to be captured and held by the precise coloring of an apple can be the stuff of poetry, even though it prolongs the shopping trip and provokes stares from busy passersby. In chapter 3 we will explore further the nature of the information contained in conceptual schemas.

Complex Schemas

Our general knowledge seems to go considerably beyond concepts for discrete objects and events, such as *apple* or *give*. Some examples that have been investigated in cognitive psychology and AI are (1) schemas, or *frames*, for complex visual scenes, such as what a room looks like (Minsky 1977), (2) schemas, or *scripts*, for complex activities, such as going to a restaurant (Schank and Abelson 1977), and (3) schemas for people's personalities, including one's own self (Cantor and Mischel 1979; Markus 1980). We will look at research on scripts, person schemas, and the self-schema.

Scripts Cognitive scientists use the term *script* or *activity schema* to refer to a declarative knowledge structure that captures general information about a routine series of events or a recurrent type of social event, such as eating in a restaurant or visiting the doctor. Schank and Abelson (1977) and other script theorists have hypothesized that scripts contain the following sorts of information: an identifying name or theme (such as *eating in a restaurant*), typical roles (*customer*, *waiter*, *cook*, . . .), entry conditions

(*customer is hungry and has money*), a sequence of goal-directed scenes (*entering and getting a table, ordering, eating, paying bill and leaving*), a sequence of actions within each scene (in the ordering scene *customer gets menu, customer reads menu, customer decides on order, customer gives order to waiter, . . .*).

Pursuing our strategy of building schemas into propositional networks, it is clear that the kind of information listed above can be captured as a set of propositions connected to a script or activity node. Some theorists have felt that assumptions must be added to propositional network theory to handle the seemingly highly organized structure of scripts. First, scripts are difficult to encode as a tangle of propositions. The script for eating in a restaurant contains a sequence of subactivities (entering, finding a table, and so on) that must be evoked in the correct order. And each subactivity (such as the ordering scene) has, in a sense, its own subscript. Furthermore, introspection suggests that entire scripts seem to be activated as units without interference from stray propositions that happen to be associated with parts of the script. Second, if scripts have any real validity as cognitive structures, we ought to be able to say more about these structures in general than that they are a cluster of interrelated propositions. That is, the scripts for different activities ought to have the same general structure: a theme, roles, entry conditions, a sequence of scenes, and so forth.

On the other hand, some theorists have pointed out that if scripts are hypothesized to be large, rigidly organized structures, it is hard to explain how they are learned from experience and how they can be flexibly retrieved and employed in situations that depart from the standard. Discussions of the theoretical issues raised by complex knowledge structures can be found in Rumelhart and Ortony 1976, Schank 1982, Anderson 1983, and Kolodner 1983. We will also return to these theoretical issues in chapter 4, since the research methods of AI are an excellent means for exploring the properties of various proposals for knowledge structure.

Without attempting to resolve these issues here, we will assume that scripts and other complex schemas can be effectively activated as units and that they exert strong influences on the way that information in the current situation is processed. We will also assume that there can be associations between complex schemas and representations of particular events that were categorized under the schema. For example, just as there can be associations between the conceptual schema for *apple* and knowledge about particular apples, so there can be associations between the restaurant script and knowledge about particular visits to restaurants.

Experimental Evidence for Scripts Bower, Black, and Turner (1979) investigated the influence of scripts on information processing by having subjects read and later try to remember brief stories about events that were instances of scripts. For example:

The Doctor

John was feeling bad today so he decided to go see the family doctor. He checked in with the doctor's receptionist, and then looked through several medical magazines that were on the table by his chair. Finally the nurse came and asked him to take off his clothes. The doctor was very nice to him. He eventually prescribed some pills for John. Then John left the doctor's office and headed home.

Although this story is easily understandable, it does not explicitly mention a number of the typical elements of the script for a visit to a doctor, such as arriving at and entering the office, or getting weighed and having one's blood pressure checked. Script

theorists hypothesize that when we hear or read about a scripted event, our knowledge of the entire script is activated, which allows us to fill in, or infer, the scenes and actions that are not explicitly mentioned. Because of this ability the story above could be told in an entirely different way, perhaps beginning as follows:

John was feeling bad today. It seemed like forever before he finally arrived at the doctor's office. He looked around at the various health posters on the wall. Finally, the nurse came and checked his blood pressure and weight. . . .

Bower, Black, and Turner hypothesized that, when trying to remember stories, subjects would tend to confuse material that they had actually read with material that they had filled in on the basis of their script knowledge. Their prediction was borne out in two experiments. In a *recall* experiment subjects read eighteen brief stories like the one just given about different people in different scripted settings and later tried to write down exactly what they had read. Subjects showed a significant tendency to recall actions that were part of the script but were not stated in the story. For example, in recalling "The Doctor," the subject might incorrectly recall reading that the nurse checked John's blood pressure and weight. The subjects showed a much smaller tendency to incorrectly recall actions that were not part of relevant scripts. A second, *recognition* memory experiment confirmed this tendency. In this experiment a different group of subjects read the same eighteen stories. They were then presented with a series of sentences and asked to rate on a 7-point scale how sure they were that they had read the sentence, from 1 ("very sure I did not read it") to 7 ("very sure I did read it"). Some of the sentences were from the stories, some were script actions that were not stated in the stories, and some were plausible actions that were neither stated in the stories nor typical script events. Subjects showed a significantly greater tendency to falsely recognize unstated script events than unstated nonscript events. These memory experiments suggest that script knowledge is used in understanding stories and that the activation of a script and its use to fill in gaps in the story leave memory traces that can become confused with the memory traces for what was actually read or heard. We will return to the study of the influences of schematization on memory in chapter 3.

Person Schemas Psychologists who study personality and social behavior have proposed that schemas exist for types of people. For example, it is plausible that we have schematic concepts for the extroverted person or the mature person. Such schemas might influence our impressions of people and our predictions of how they will behave in various situations. Cantor and Mischel (1979) showed that people agree on the typical characteristics of different personality types and that person schemas influence the initial perception of and later memory for other people. For example, when subjects read a description of an extrovert that explicitly mentioned features that were only moderately typical of extroversion, the subjects apparently inferred or "filled in" the more typical features that were not mentioned. In a later recognition memory test they tended to incorrectly remember the highly typical features as having been mentioned in the original description. Cantor, Mischel, and Schwartz (1982) demonstrated an interaction between schemas for social situations (for instance, a party) and schemas for personality types (for instance, extrovert). Subjects' predictions about how a person would behave were a joint product of the situation and person schemas.

The Self-Schema One's knowledge of oneself probably occupies a special place in the store of knowledge about people, and psychologists have proposed that the self-concept is a particularly elaborate and influential schema. Markus (1980), for example,

has proposed that the self-schema exerts powerful effects on the person's perception, memory, and thought. Information that is consistent with the schema is processed more efficiently and better remembered. Incoming information is hypothesized to be evaluated relative to the schema and resisted if it is inconsistent. Plans and predictions of future behavior are crucially influenced by the schema.

Markus hypothesized that some characteristics will be part of a person's self-schema and that others will not. For example, a person may conceptualize himself or herself in terms of a cluster of attributes that were traditionally considered masculine, such as competitive and ambitious, or in terms of a cluster of attributes that were traditionally considered feminine, such as gentle and sensitive (Markus et al. 1982). Markus termed such people *masculine* and *feminine schematics*. She also hypothesized that there are people who are aschematic with respect to traditional notions of gender in the sense that they do not conceptualize themselves in terms of these traditional clusters of attributes. Using a personality test called the Bem Sex Role Inventory (Bem 1974), she was able to classify her subjects as masculine schematics, feminine schematics, or aschematic with respect to traditional gender schemas. She was then able to show that subjects' behavior on a variety of information-processing tasks was influenced by their gender schematization. For example, Markus et al. (1982) collected response times in a property verification task concerning the self. In this task subjects were presented with adjectives and asked to judge as quickly as possible whether those adjectives were self-descriptive. Masculine schematic subjects answered most quickly for traditionally prototypical masculine adjectives, and feminine schematics answered most quickly for feminine adjectives, whereas aschematics showed the same response times for masculine, feminine, and neutral (for instance, "friendly") adjectives.

Propositions, Schemas, and the Architecture of Cognition

Now that we have examined some evidence for propositions and schemas of various kinds, we can return to the claim that they are part of the cognitive architecture. The claim that propositional representation is part of the cognitive architecture is central to the classical view. Notice that it is the *facility* for propositional representation, not any particular proposition, that is hypothesized to be built into the architecture of cognition. Propositions are simple, and therefore easy to implement physically, but they have the productive syntactic structure required by the classical view. They also have considerable power to represent the world (an issue taken up in more detail in chapter 10). The usefulness of a long-term memory full of propositions depends on distal access. There must be an efficient mechanism for retrieving particular propositions from memory when they are relevant or needed for ongoing cognition. Spreading activation in an associative network has proven to be a fruitful hypothesis about the retrieval mechanism, although other mechanisms have been proposed. The potential hypotheses about the basic scheme for long-term memory retrieval include any efficient *content-addressing* mechanism, that is, any mechanism that takes propositions or fragments of propositions as inputs and locates other propositions in memory with overlapping content.

The classical view is also strongly associated with the claim that mechanisms for organizing propositional knowledge into abstract schemas are built into the biology of cognition. The representational format of schemas and the nature of the abstraction process—that is, the process by which concepts, scripts, and so on, are learned—are matters of current debate and research, some of which will be described in chapter 3.

Cognition requires some mechanism for abstraction, or for assimilating information to known concepts or categories. The idea that conceptual knowledge is structured so that it can be retrieved without interference and used efficiently also makes a great deal of sense. Augmenting propositional representation to do these jobs is a natural elaboration of the classical view.

2.6 Cognitive Processes, Working Memory, and Attention

Before continuing our exploration of cognitive representation, we must consider how representations are processed. So far, propositional-schematic representations have been proposed as a medium for the storage of information in human *long-term memory*. Associative activation has been proposed as a scheme for retrieving that information. In terms of the mechanics of the architecture an item of information counts as retrieved if its activation is above some threshold. (Alternatively we could say the probability of retrieval is some continuous, increasing function of activation giving every proposition in memory some nonzero chance of retrieval at any given moment.) Propositions that are retrieved are available for further processing or for motor output. Earlier, we also associated retrieval with availability to consciousness.

The retrieval mechanism alone obviously does not suffice to account for our thoughts or actions. It delivers an associative stream of information. There have to be further mechanisms that assess retrieved information for relevance to current goals and that can weld pieces of retrieved information into coherent streams of thought that lead to conclusions, decisions, or solutions to problems.

Goal Orientation and Flexibility of Cognitive Processes

In addition to the associative retrieval process, then, we must hypothesize the existence of other processes that serve to transform the information that is active at any given moment. The active information will typically include not only information from long-term memory but also information arising from sensory systems and information about current goals. The processes required are reasoning, planning, or problem-solving processes that can take the available information as inputs and produce conclusions, decisions, or action commands as outputs.

Suppose one had a goal to spend some time reading a book, and the current visual input yielded the information that it was too dark to read. A simple reasoning process might combine these two pieces of information and yield a *subgoal* to increase the level of illumination. This goal might trigger a visual search for a source of electric light, producing visual information that there is a lamp near one's chair, which in turn might lead to a goal of turning on the lamp, triggering a visual search for the switch, yielding visual information that in turn triggers the motor action of reaching for the switch. Suppose that at that point the lamp does not go on. This information might trigger a suspension of the top-level reading goal, which might lead to the retrieval of information that a paper on the material to be read is due in the very near future, which might lead to a restoration of the reading goal. The restored goal might lead to the retrieval of a schema concerning electrical appliances that includes the fact that they usually have to be plugged into an electrical outlet. This fact might trigger the construction of a new subgoal to check whether the lamp is plugged in. This goal might trigger a visual search for the disposition of the plug at the end of the cord or perhaps the production of a request to a friend in the room, "Could you check to see if this lamp