THE MIND'S NEW SCIENCE

A History of the Cognitive Revolution

◆ ◆

Howard Gardner

WITH A NEW EPILOGUE BY THE AUTHOR:
Cognitive Science After 1984

THE MIND'S NEW SCIENCE

A History of the Cognitive Revolution

HOWARD GARDNER

With a New Epilogue by the Author:

Cognitive Science After 1984

BASIC BOOKS
A Member of the Perseus Books Group
New York

Quotes on pp. 66 and 88 from H. Putnam, *Mind, Language, and Reality: Philosophical Papers, vol. 2,* 1975. Reprinted with permission, Cambridge University Press.
Quotes on pp. 20, 24, 132, and 295 from N. Wiener, *Cybernetics, or Control and Communication in the Animal and the Machine,* 2nd ed., 1961/1948. Reprinted with permission, MIT Press.
Quote on p. 244 from Jonathan Miller, *States of Mind,* 1983. Reprinted with permission, Pantheon Books, a Division of Random House, Inc.
Quotes on pp. 72, 73, 74, 75, and 85 from Richard Rorty, *Philosophy and the Mirror of Nature.* Copyright © 1979 by Princeton University Press. Excerpts reprinted with permission of Princeton University Press.
Quote on p. 70 from Jerome Bruner, *In Search of Mind* (in press). Reprinted with permission of the author.

Library of Congress Cataloging in Publication Data

Gardner, Howard.
The mind's new science.

Bibliography: p. 401.
Includes index.
1. Cognition—History. 2. Cognition—Research—Methodology—History. 3. Artificial intelligence—History. I. Title.
BF311.G339 1985 153 85-47555
ISBN 0-465-04634-7 (cloth)
ISBN-10: 0-465-04635-5 ISBN-13: 978-0-465-04635-5 (paper)

Printed in the United States of America
Designed by Vincent Torre

CONTENTS

PART II

THE COGNITIVE SCIENCES: A HISTORICAL PERSPECTIVE

PART III
TOWARD AN INTEGRATED COGNITIVE SCIENCE: PRESENT EFFORTS, FUTURE PROSPECTS

PREFACE

In the mid-1970s, I began to hear the term *cognitive science.* As a psychologist interested in cognitive matters, I naturally became curious about the methods and scope of this new science. When I was unable to find anything systematic written on the subject, and inquiries to colleagues left me confused, I decided to probe further. Some immersion in the writings of self-proclaimed cognitive scientists convinced me that cognitive science was deeply rooted in philosophy and therefore, in a sense, had a long history. At the same time, the field was so new that its leading figures were all alive, and some of them were still quite young.

I decided that it would be useful and rewarding to undertake a study in which I would rely heavily on the testimony of those scholars who had founded the field as well as those who were at present its most active workers. But in lieu of an oral history or a journalistic account of current laboratory work (both of which subsequently were undertaken by other authors), I decided to make a comprehensive investigation of cognitive science in which I could include the long view—the philosophical origins, the histories of each of the respective fields, the current work that appears most central, and my own assessment of the prospects for this ambitious field.

It had not escaped my attention that the Alfred P. Sloan Foundation was a major supporter of work in the cognitive sciences. I therefore approached its program officer Kenneth Klivington about the possibility of writing a history of cognitive science. To my delight, the Foundation proved receptive, and I began my formal study at the beginning of 1981. I want to express my gratitude to the entire administration of the Sloan Foundation, and to its two responsible program officers, Kenneth Klivington and Eric Wanner, who were totally supportive of my efforts to carry through this somewhat risky undertaking.

In the course of my study, I interviewed formally, or conducted informal discussions with, dozens of cognitive scientists in this country and abroad. As far as I can recall, no scientist whom I approached denied me an interview, and most—even those who expressed skepticism about cognitive science—were gracious and informative. I regret that I had to stop interviewing and begin writing after a time, and I regret even more that I ultimately was not able to discuss in print the work of many of those from whom I learned much. Unfortunately, if I had included even half of the work worthy of review, this book would be several times longer than it is.

I want to mention first and thank the many individuals who willingly discussed their work and the field of cognitive science with me. (I also must apologize to those whom I have inadvertently omitted from this list.) I am indebted to Jonathan Adler, Allan Allport, John Anderson, Dana Ballard, Jon Barwise, Elizabeth Bates, Brent Berlin, Ned Block, Daniel Bobrow, Margaret Boden, Stanley Brandes, Joan Bresnan, John Seely Brown, Roger Brown, Jerome Bruner, Peter Bryant, Alfonso Caramazza, Noam Chomsky, Gillian Cohen, Michael Cole, Roy D'Andrade, Daniel Dennett, Hubert Dreyfus, Jerome Feldman, Charles Fillmore, Jerry Fodor, Michael Gazzaniga, Clifford Geertz, my late and beloved mentor Norman Geschwind, Samuel Glucksberg, Nelson Goodman, Charles Gross, Patrick Hayes, Geoffrey Hinton, Stephen Isard, Philip Johnson-Laird, Ronald Kaplan, Paul Kay, Samuel Jay Keyser, Stephen Kosslyn, George Lakoff, Jean Lave, Jerome Lettvin, Robert LeVine, Claude Lévi-Strauss, Christopher Longuet-Higgins, John McCarthy, Jay McClelland, Jean Mandler, Alexander Marshack, John Marshall, Jacques Mehler, Susanna Millar, George Miller, Marvin Minsky, Julius Moravcsik, John Morton, Ulric Neisser, Freda Newcombe, Allen Newell, Donald Norman, Daniel Osherson, Domenico Parisi, Stanley Peters, Michael Posner, Karl Pribram, Hilary Putnam, Raj Reddy, Richard Rorty, Eleanor Rosch, David Rumelhart, Roger Schank, Israel Scheffler, John Searle, Robert Siegler, Herbert Simon, Aaron Sloman, Brian Cantwell Smith, Stuart Sutherland, Leonard Talmy, Sheldon Wagner, Terry Winograd, and Edgar Zurif.

Several friends and colleagues were good enough to read and comment critically on one or more of the drafts of this book. I am considerably in their debt. I wish to thank Margaret Boden, Hiram Brownell, Daniel Dennett, Martha Farah, Josef Grodzinsky, Jerome Kagan, Benny Shanon, Eric Wanner, several anonymous reviewers, and my wife, Ellen Winner, for their useful comments, criticisms, and words of encouragement. I know that I benefited greatly from their feedback; I fear that remaining errors and infelicities are my own responsibility.

Over the several years in which this book was in preparation, I was fortunate enough to have the help of Linda Levine, Susan McConnell, Christine Meyer, and Claudia Strauss, who served as research assistants.

Mara Krechevsky, my current research assistant, has been invaluable in helping me to bring the manuscript to publication. In addition, she has made many substantive contributions to the manuscript. I thank Connie Wolf at Harvard and Carmella Loffredo at the Sloan Foundation for their help. The manuscript in its various guises was ably typed and word-processed by Dolly Appel, Damaris Chapin, Isabel Eccles, Nan Kortz, and Laura Stephens-Swannie. I am sure they would agree with the sentiment expressed by Samuel Johnson with respect to *Paradise Lost:* "No man could wish it longer."

As with my last three books, I have been fortunate to have the support of many individuals at Basic Books. On the editorial side I am tremendously grateful to Judith Greissman, Jane Isay, and Martin Kessler for their thoughtful reactions to earlier versions of this manuscript. Linda Carbone performed ably as the project editor; and Phoebe Hoss, as development editor, helped me to deal with many expositional problems and also displayed an uncanny sense of where I (and, at times, where cognitive science) had fallen short. In another life, she is at risk of becoming a cognitivist herself.

My greatest pleasure is to have the opportunity to dedicate this book to my parents.

HOWARD GARDNER
Cambridge, Massachusetts
April 1985

PART I

THE COGNITIVE REVOLUTION

1

Introduction: What the *Meno* Wrought

One thing I would fight for to the end, both in word and deed if I were able—that if we believed that we must try to find out what is not known, we should be better and braver and less idle than if we believed that what we do not know it is impossible to find out and that we need not even try.

—Socrates, The *Meno*

The safest general characterization of the European philosophical tradition is that it consists in a series of footnotes to Plato.

—Alfred North Whitehead

The Greek Agenda

In the *Meno,* a Platonic dialogue, Socrates persistently questions a young slave about his knowledge of geometry. At first the slave appears quite knowledgeable, readily asserting that a square composed of sides two feet in length contains four square feet. But when, in response to a problem posed by Socrates, the slave indicates that a figure of eight square feet contains sides four feet long, Socrates demonstrates that the boy is thoroughly confused and does not realize that the length of the side must be the square root of eight.

The centerpiece of the dialogue features many questions and responses in the approved Socratic manner. Through this interchange, the philosopher ultimately succeeds in drawing out from the boy the knowledge that a square with a four-foot side would actually be sixteen square feet—that is, twice as great an area than he had supposed; and the knowledge that one can, by geometric maneuvers, inscribe a square that is actually eight square feet within this larger square. In so doing, Socrates has demonstrated to his satisfaction, and to the satisfaction of the slave's master, Menon, that the youth possesses within him all of the knowledge necessary to compute the various geometrical relationships in question.

At issue in this Platonic dialogue was far more than an exploration of the extent of knowledge possessed by a single slave boy. Here, for perhaps the first time in human intellectual history, was an extended rumination on the nature of knowledge: where does it come from, what does it consist of, how is it represented in the human mind? And, for good measure, there was also proposed a specific—if ultimately highly controversial—theory of human knowledge.

According to Plato (and, presumably, Socrates as well), the domain of knowledge par excellence inhered in mathematics and the exact sciences it had spawned. Indeed, the purest forms of knowledge were idealized forms or archetypes which can merely be glimpsed in mundane reality. An understanding of geometrical matters—indeed, of all matters of genuine knowledge—was already implanted in the human soul at birth. The task in instruction, as demonstrated in the dialogue of the *Meno,* was simply to bring this innate knowledge to conscious awareness.

The Greeks' interest in the nature of knowledge, no less than their particular contentious theories and evocative images, continued to reverberate through the Western intellectual tradition. Aristotle's version was the principal cornerstone of the Middle Ages, when discussions about knowledge were principally the purview of theologians. Then, during the Renaissance and Enlightenment periods, philosophers continued the discussions and began to draw regularly on findings obtained in the newly emerging empirical sciences. Such thinkers as Descartes, Locke, and Kant dealt comfortably with theoretical and empirical issues concerning knowledge, and the Neapolitan scholar Giambattista Vico even christened a New Science *(Scienza Nova)* to deal with these and related matters. By the end of the nineteenth century, there had been a proliferation of new sciences and philosophical specialties, several of which purported to deal with the nature of the human mind.

Today, armed with tools and concepts unimaginable even a century ago, a new cadre of thinkers called cognitive scientists has been investigating many of the same issues that first possessed the Greeks some twenty-five hundred years ago. Like their earlier counterparts, cognitive scientists today ask what it means to know something and to have accurate beliefs, or to be ignorant or mistaken. They seek to understand what is known—the objects and subjects in the external world—and the person who knows—his* perceptual apparatus, mechanisms of learning, memory, and rationality. They ponder the sources of knowledge: where does it come from, how is it stored and tapped, how might it be lost? They are curious about the differences among individuals: who learns early or with difficulty; what can be known by the child, the inhabitant of a preliterate society, an individual who has suffered brain damage, or a mature scientist?

Further, cognitive scientists, again as did the Greeks, conjecture about the various vehicles of knowledge: what is a form, an image, a concept, a word; and how do these "modes of representation" relate to one another? They wonder about the priorities of specific sense organs as against a central "general understanding" or "common sense." They reflect on language, noting the power and traps entailed in the use of words and their possible predominant influence over thoughts and beliefs. And they speculate at length on the nature of the very activity of knowing: why do we want to know, what are the constraints on knowing, and what are the limits of scientific knowledge about human knowing?

This "new science," thus, reaches back to the Greeks in the commitment of its members to unraveling the nature of human knowledge. At the same time, however, it is radically new. Proceeding well beyond armchair speculation, cognitive scientists are fully wedded to the use of empirical methods for testing their theories and their hypotheses, of making them susceptible to disconfirmation. Their guiding questions are not just a rehash of the Greek agenda: new disciplines, like artificial intelligence, have arisen; and new questions, like the potential of man-made devices to think, stimulate research. Moreover, cognitive scientists embrace the most recent scientific and technological breakthroughs in a variety of disciplines. Most central to their undertaking is the computer—that creation of the mid-twentieth century that holds promise for changing our conceptions of the world in which we live and our picture of the human mind.

Definition and Scope of Cognitive Science

In the course of proposing and founding a new field of knowledge, many individuals will formulate their own definitions. Indeed, since the term *cognitive science* first began to be bandied about in the early 1970s, dozens of scientists have attempted to define the nature and scope of the

*For ease of exposition, the pronoun *he* is used in its generic sense throughout this book.

field (see, for example, Bruner 1983; Collins 1977; Mandler 1981; Miller 1979; Norman 1980; Rumelhart 1982). It therefore becomes important for me at the outset to state what I take cognitive science to be.

I define cognitive science as a contemporary, empirically based effort to answer long-standing epistemological questions—particularly those concerned with the nature of knowledge, its components, its sources, its development, and its deployment. Though the term *cognitive science* is sometimes extended to include all forms of knowledge—animate as well as inanimate, human as well as nonhuman—I apply the term chiefly to efforts to explain human knowledge. I am interested in whether questions that intrigued our philosophical ancestors can be decisively answered, instructively reformulated, or permanently scuttled. Today cognitive science holds the key to whether they can be.

Of the various features or aspects generally associated with cognitive-scientific efforts, I consider five to be of paramount importance. Not every cognitive scientist embraces every feature, of course; but these features can be considered symptomatic of the cognitive-scientific enterprise. When all or most are present, one can assume that one is dealing with cognitive science; when few, if any, are present, one has fallen outside my definition of cognitive science. These features will be introduced more formally at the end of chapter 3 and will be revisited repeatedly throughout the book, but it is important to make an initial acquaintance with them at this point.

First of all, there is the belief that, in talking about human cognitive activities, it is necessary to speak about mental representations and to posit a level of analysis wholly separate from the biological or neurological, on the one hand, and the sociological or cultural, on the other.

Second, there is the faith that central to any understanding of the human mind is the electronic computer. Not only are computers indispensable for carrying out studies of various sorts, but, more crucially, the computer also serves as the most viable model of how the human mind functions.

While the first two features incorporate the central beliefs of current cognitive science, the latter three concern methodological or strategic characteristics. The third feature of cognitive science is the deliberate decision to de-emphasize certain factors which may be important for cognitive functioning but whose inclusion at this point would unnecessarily complicate the cognitive-scientific enterprise. These factors include the influence of affective factors or emotions, the contribution of historical and cultural factors, and the role of the background context in which particular actions or thoughts occur.

As a fourth feature, cognitive scientists harbor the faith that much is to be gained from interdisciplinary studies. At present most cognitive scientists are drawn from the ranks of specific disciplines—in particular, philosophy, psychology, artificial intelligence, linguistics, anthropology, and neuroscience (I shall refer to these disciplines severally as the "cognitive sciences"). The hope is that some day the boundaries between these disciplines may become attenuated or perhaps disappear altogether, yielding a single, unified cognitive science.

A fifth and somewhat more controversial feature is the claim that a key ingredient in contemporary cognitive science is the agenda of issues, and the set of concerns, which have long exercised epistemologists in the Western philosophical tradition. To my mind, it is virtually unthinkable that cognitive science would exist, let alone assume its current form, had there not been a philosophical tradition dating back to the time of the Greeks.

Purpose and Plan of This Book

I have chosen to write a book on cognitive science because I consider this area to be the most exciting new line of inquiry undertaken by scientists in the past few decades. Whether it will ultimately achieve all of its objectives, no one can say at this point; but this seems an opportune time to present a history and a current assessment. For contemporaries present during the opening decades of cognitive science, I hope to convey something of the enthusiasm I have noted, the difficulties that are being confronted, and the nature of the research enterprises in which investigators are presently engaged.

My history has two components. The first consists of the various interdisciplinary conversations and projects that took place in this century —both those preceding and those surrounding the unofficial launching of cognitive science in the mid-1950s. I relate the founding of cognitive science in the next two chapters of the book. The second component—spanning chapters 4 through 9—consists of brief targeted histories of each of the six aforementioned fields of cognitive science. (Other disciplines, such as sociology or economics, might have been added; the "borderline" disciplines of anthropology and neuroscience might have been eliminated; but I believe that the major points about cognitive science are made effectively by these six fields.) In my view, a brief targeted history of each of the several cognitive sciences serves as an optimal introduction to the principal issues of today, to the ways in which they are currently approached and explored, and to the lines of work likely to be undertaken in the future.

I have built each historical chapter around one or two major themes, which have been selected to convey a feeling for the kinds of issues that have recurred and the kinds of approaches that are especially central within a particular field. For example, in philosophy I trace the perennial dispute between those of a rationalist persuasion (who view the mind as actively organizing experiences on the basis of pre-existing schemes); and those of an empiricist bent (who treat mental processes as a reflection of information obtained from the environment). In anthropology I survey various attempts over the years to compare the thought of primitive peoples with that exhibited by typical individuals in modern Western society. Approaching these same fields from a methodological point of view, I raise the questions whether philosophy will eventually come to be supplanted by an empirically based cognitive science, and whether anthropology can (or even should) ever transcend the individual case study.

Of course, such organizing themes can only scratch the surface of the complex territory that underlies any scientific discipline. Still I hope that through such themes I can convey how a linguist views an issue, what a psychologist deems a problem (and a solution), which conceptions of process obtain in neuroscience and artificial intelligence. Only through such an immersion in the daily (and yearly) concerns of a cognitive scientist drawn from a particular discipline can one appreciate the possibilities —and the difficulties—that arise when workers from different fields collaborate in cognitive-scientific research. In the end I will in each case take stock and indicate where things stand with reference to the principal lines of contention in a particular cognitive science—a discussion that will, in turn, suggest some of the principal factors that have stimulated cognitive scientists to join forces.

While each of the histories stands on its own, their juxtaposition points up fascinating and difficult-to-anticipate parallels. Scientific fields hardly develop in a vacuum: such disparate factors as the dissemination of Darwin's pivotal writings, the outbreak of wars, the rise of great universities have had reverberations—and sometimes cataclysmic ones—across apparently remote fields, which may well have had little direct contact with one another. For the most part, I shall simply allow these parallels to emerge, but at the beginning of part III I shall specify certain historical forces that seem to have exerted influence across a range of cognitive sciences.

Having taken the measure of the individual cognitive sciences, I turn in the third part of the book to review ongoing work that is quintessentially cognitive-scientific. Thus, in chapters 10 to 13, the focus shifts from work within a traditional discipline to those lines of research that stand most squarely at the intersection of a number of disciplines and therefore can be considered prototypical of a single, unified cognitive science. I have sought to identify work that is of the highest quality: if cognitive science is to be assessed as an intellectual enterprise, it ought to be judged by the most outstanding instances.

There is a common structure to these four essays on current cognitive-scientific work. Consistent with my claim that cognitive science seeks to elucidate basic philosophical questions, each chapter begins with a perennial epistemological issue. For example, in chapter 10, I describe work on how we perceive the world; in chapter 13, I review competing claims on the extent of human rationality. Across chapters 10 to 13, there is a progression from those issues that seem most circumscribed to those that are most global. Not surprisingly, the most confident answers exist for the delimited questions, while the global topics remain ringed by unresolved questions.

My personal reflections on cognitive science are reserved for the final chapter. There I revisit the major themes of cognitive science in light of the histories sketched and the interdisciplinary work reviewed. I also discuss two themes that emerge from the inquiry and that will be introduced at greater length in chapter 3: the computational paradox and the cognitive challenge. In my view, the future of cognitive science rests on how the computational paradox is resolved and on how the cognitive challenge is met.

One might say that cognitive science has a very long past but a relatively short history. The reason is that its roots go back to classical times, but it has emerged as a recognized pursuit only in the last few decades. Indeed, it seems fair to maintain that the various components that gave rise to cognitive science were all present in the early part of the century, and the actual birthdate occurred shortly after mid-century. Just why cognitive science arose when it did in the form it did will constitute my story in the remainder of part I.

2

Laying the Foundation for Cognitive Science

The Hixon Symposium and the Challenge to Behaviorism

In September of 1948 on the campus of the California Institute of Technology, a group of eminent scientists representing several disciplines met for a conference on "Cerebral Mechanisms in Behavior," sponsored by the Hixon Fund (Jeffress 1951). This conference had been designed to facilitate discussions about a classic issue: the way in which the nervous system controls behavior. And yet the discussions ranged far more widely than the official topic had implied. For example, the opening speaker, mathematician John von Neumann, forged a striking comparison between the electronic computer (then a discovery so new that it smacked of science fiction) and the brain (which had been around for a while). The next speaker, mathematician and neurophysiologist Warren McCulloch, used his provocative title ("Why the Mind Is in the Head") to launch a far-ranging discussion on how the brain processes information—like von Neumann, he wanted to exploit certain parallels between the nervous system and "logical devices" in order to figure out why we perceive the world the way we do.

Less steeped in the latest technological innovations but more versed in the problems of explaining human behavior, the next speaker, psychologist Karl Lashley, gave the most iconoclastic and most memorable address. Speaking on "The Problem of Serial Order in Behavior," he challenged the doctrine (or dogma) that had dominated psychological analysis for the past several decades and laid out a whole new agenda for research. In the terms of my own discussion, Lashley identified some of the major components needed for a cognitive science, even as he castigated those forces that had prevented its emergence before this time.

In order to appreciate the importance of Lashley's remarks, it is necessary to consider the scientific climate in which he (and his numerous colleagues interested in human psychology) had been operating during the past few decades. At the turn of the century, in the wake of the founding of new human sciences, investigators had been addressing the key issues of mental life: thinking, problem solving, the nature of consciousness, the unique aspects of human language and culture. These discussions had linked up with the philosophical agenda of the West, but investigators had sought to go beyond sheer speculation through the use of rigorous experimental methods.

Unfortunately the scientific method favored by most researchers at that time was introspection: self-reflection on the part of a trained observer about the nature and course of his own thought patterns. Though suggestive (indeed, often too suggestive), such introspection did not lead to that accumulation of knowledge that is critical to science. Introspectionism might have collapsed of its own weight, but, in fact, it was toppled in a more aggressive manner by a group of mostly young, mostly American scientists who became known as the "behaviorists."

The behaviorists put forth two related propositions. First of all, those researchers interested in a science of behavior ought to restrict themselves strictly to public methods of observations, which any scientist could apply and quantify. No subjective ruminations or private introspection: if a discipline were to be science, its elements should be as observable as the physicist's cloud chamber or the chemist's flask. Second, those interested in a science of behavior ought to focus exclusively on *behavior:* researchers ought assiduously to eschew such topics as mind, thinking, or imagination and such concepts as plans, desires, or intentions. Nor ought they to countenance hypothetical mental constructs like symbols, ideas, schemas, or other possible forms of mental representation. Such constructs, never adequately clarified by earlier philosophers, had gotten the introspectionists into hot water. According to behaviorists, all psychological activity can be adequately explained without resorting to these mysterious mentalistic entities.

A strong component of the behaviorist canon was the belief in the supremacy and determining power of the environment. Rather than individuals acting as they do because of their own ideas and intentions, or because their cognitive apparatuses embody certain autonomous structuring tendencies, individuals were seen as passive reflectors of various forces and factors in their environment. An elaborate explanatory apparatus detailing principles of conditioning and reinforcement was constructed in order to explain just how such learning and shaping of particular behavior might come about. It was believed that the science of behavior, as fashioned by scholars such as Ivan Pavlov, B. F. Skinner, E. L. Thorndike, and J. B. Watson, could account for anything an individual might do, as well as the circumstances under which one might do it. (What one thinks was considered irrelevant from this perspective—unless thought was simply redefined as covert behavior.) Just as mechanics had explained the laws of the physical world, mechanistic models built on the reflex arc could explain human activity.

Behaviorism spoke to many needs in the scientific community, including some that were quite legitimate: discomfort with the acceptance of introspective evidence on face value, without any means of scientific control or any possibility for refutation; dissatisfaction with vague and global concepts like *will* or *purpose* and the desire to explain human behavior using the same constructs that were applied (with apparently great success) to animal behavior. Indeed, in the wake of the troubles that had arisen from reliance on introspectionism (troubles that are spelled out in chapter 4), behaviorism seemed like a breath of fresh air during the opening decades of the century. Little wonder that it caught on quickly and captured the best minds of a generation of workers.

Yet, in retrospect, the price paid by strict adherence to behaviorism was far too dear. So long as behaviorism held sway—that is, during the 1920s, 1930s, and 1940s—questions about the nature of human language, planning, problem solving, imagination, and the like could only be approached stealthily and with difficulty, if they were tolerated at all. Lashley's paper crystallized a growing awareness on the part of thoughtful scientists that adherence to behaviorist canons was making a scientific study of mind impossible.

Lashley realized that before new insights about the brain, or about computers, could be brought to bear in the psychological sciences, it would be necessary to confront behaviorism directly. Therefore, in his opening remarks, Lashley voiced his conviction that any theory of human activity would have to account for complexly organized behaviors like playing tennis, performing on a musical instrument, and—above all—speaking. He commented, "The problems raised by the organization of language seem to me to be characteristic of almost all other cerebral activity" (quoted in Jeffress 1951, p. 121). In this one sentence, Lashley put at the very center of human psychology a topic that had been relegated to obscurity by his behaviorist colleagues. At the same time, he added, the dominant theoretical explanatory framework in neurophysiology no less than in psychology —that of simple associative chains between a stimulus and a response— could not possibly account for any of this serially ordered behavior. The reason is that these action sequences unfold with such rapidity that there is no way in which the next step in the chain can be based upon the previous one: when one plays an arpeggio, for instance, there is simply no time for feedback, no time for the next tone to depend upon or in any way to reflect the course of the preceding one. Similarly, the kinds of error made by individuals—for example, slips of the tongue—often include anticipation of words that are to occur only much later in a sequence. Again, these phenomena defy explanations in terms of linear "A evokes B" chains.

According to Lashley, these behavioral sequences have to be planned and organized in advance. The organization is best thought of as hierarchical: there are the broadest overall plans, within which increasingly fine-grained sequences of actions are orchestrated. Thus, for instance, in the case of speech, the highest nodes of the hierarchy involve the overall intention prompting the utterance, while the choice of syntax and the actual production of sounds occupy lower nodes of the hierarchy. The nervous system contains an overall plan or structure, within which individual response units can—indeed, have to—be slotted, independent of specific feedback from the environment. Rather than behavior being consequent upon environmental promptings, central brain processes actually precede and dictate the ways in which an organism carries out complex behavior. Or, to put it simply, Lashley concluded that the form precedes and determines specific behavior: rather than being imposed from without, organization emanates from within the organism.

Even as he defied the standard behavioral analysis of the time, Lashley was also challenging two major dogmas of neurobehavioral analysis: the belief that the nervous system is in a state of inactivity most of the time, and the belief that isolated reflexes are activated only when specific forms of stimulation make their appearance. Lashley's nervous system consisted of always active, hierarchically organized units, with control emanating from the center rather than from peripheral stimulation. As he put it, "Attempts to express cerebral function in terms of the concepts of the reflex arc, or of associated chains of neurons, seem to me doomed to failure because they start with the assumption of a static nervous system. Every bit of evidence available indicated a dynamic, constantly active system, or, rather, a composite of many interacting systems" (quoted in Jeffress 1951, p. 135).

In the topics he chose to address, and in the ways in which he addressed them, Lashley was adopting a radical position. Scientists concerned with human behavior had been reluctant to investigate human language, because of its complexity and its relative "invisibility" as a form of behavior; and when they did treat language, they typically sought analogies to simpler forms (like running a maze or pecking in a cage) in simpler organisms (like rats or pigeons). Not only did Lashley focus on language, but he reveled in its complexity and insisted that other motoric activities were equally intricate.

Ordinarily, a scientist who challenges established wisdom is in for a rough time. It is rare, at a scientific meeting, for major scholars (an ambitious and often jealous lot) to pay homage to a colleague. But from comments by those attending the Hixon Symposium, it seemed clear that Lashley's colleagues were deeply impressed by the originality and brilliance of this presentation—coming from a scientist closely associated with the behaviorist tradition. Lashley himself declared, "I have been rather embarrassed by some of the flattering remarks made today" (quoted in Jeffress 1951, p. 144). It is no exaggeration to suggest that entrenched modes of explanation were beginning to topple and that a whole new agenda was confronting the biological and behavioral communities.

A Critical Moment in Scientific History

The scholars in attendance at the Hixon Symposium stood at a critical juncture of scientific history. They were keenly aware of the staggering advances of previous centuries in the physical sciences as well as of recent breakthroughs in the biological and neural sciences. Indeed, by the middle of the twentieth century, two major mysteries of ancient times—the nature of physical matter and the nature of living matter—were well on their way to being unraveled. At the same time, however, a third mystery that had also fascinated the ancients—the enigma of the human mind—had yet to achieve comparable clarification.

Trained (like many scholars of their time) in the humanities as well as in the sciences, the Hixon symposiasts displayed a familiarity with the kinds of epistemological issue that had first exercised the Greeks and had then formed a major part of learned conversation during the Enlightenment. They knew that, in the wake of Darwin's influential account of the origin and evolution of species, many scientists had sought to bring comparable rigor to the study of human behavior and thought. Often spurning direct ties to philosophy (which they regarded as a regressive intellectual force), these scholars at the end of the nineteenth century had launched separate scientific disciplines, like psychology, linguistics, anthropology, sociology, and various neurosciences. That these aspiring scientists of human nature had succeeded in establishing effective institutional bases within the universities could not be disputed; but the extent to which each new discipline had arrived at important truths was still being debated at mid-century. Finally, those attending the Pasadena meeting were well acquainted with the scientific program of the behaviorists. And they shared an intuition—strongly bolstered by Lashley's tightly reasoned paper—that the behaviorist answer to questions of the human mind was no answer at all.

But other factors had also impeded the proper launching of a science of cognition. Fitting comfortably with behaviorism were several philosophical schools—positivism, physicalism, and verificationism—which eschewed entities (like concepts or ideas) that could not be readily observed and reliably measured. There was also the intoxication with psychoanalysis. While many scholars were intrigued by Freud's intuitions, they felt that no scientific discipline could be constructed on the basis of clinical interviews and retrospectively constructed personal histories; moreover, they deeply resented the pretense of a field that did not leave itself susceptible to disconfirmation. Between the "hard line" credo of the Establishment behaviorists and the unbridled conjecturing of the Freudians, it was difficult to focus in a scientifically respectable way on the territory of human thought processes.

Finally, the world political situation had exerted a crippling effect on the scientific enterprise. First, the European scientific establishment had been ripped apart by the rise of totalitarianism, and then the American scientific establishment had been asked to lay aside its theoretical agenda in order to help wage the war.

While the war had been, in many ways, the worst of times, bringing on the death or disability of many talented investigators, it had also stimulated certain scientific and technological activities. Within the United States, the war effort demanded calculating machines that could "crunch" large sets of numbers very quickly. Computers soon became a reality. There were other war needs to be met as well. For instance, the mathematician Norbert Wiener was asked to devise more accurate anti-aircraft machinery. This work required "a good gun, a good projectile, and a fire-control system that enables the gunner to know the target's position, apply corrections to the gun controls, and set the fuse properly, so that it will detonate the projectile at the right instant" (quoted in Heims 1980, p. 183). While working on these problems at the Massachusetts Institute of Technology, Wiener and his associate, a young engineer named Julian Bigelow, concluded that there were important analogies between the feedback aspects of engineering devices and the homeostatic processes by which the human nervous system sustains purposive activity. These ideas of planning, purpose, and feedback, developed with mathematical precision, were directly antithetical to the behaviorist credo. War also produced many victims of gunfire; and medical practitioners who cared for brain-injured patients were being asked to evaluate which tasks could be carried out and which ones had been compromised—temporarily or permanently—by injury to the nervous system. Also, a host of more person-centered issues—ranging from the study of the effects of propaganda to the selection of men fit to lead combat units—enlisted the efforts of behavioral scientists and generated ideas on which the postwar human sciences were to build (Bruner 1944; Murray 1945; Stouffer et al. 1949). So it was in other war-torn lands, from Alan Turing and Kenneth Craik's interest in computers in England, to Alexander Luria's painstaking research with brain-injured patients in Russia during the war.

By the late 1940s, there was beginning to be a feeling abroad—one observable at Pasadena but in no way restricted to that site—that perhaps the time was ripe for a new and finally effective scientific onslaught on the human mind. Interestingly, nearly all of the work that came to fruition in the postwar era was in fact built upon prior theoretical efforts—work often dating back to the beginning of the century. But this work had sometimes been obscured by the behaviorist movement and had sometimes been transformed in unanticipated ways by the events of the war. These ideas, these key inputs to contemporary efforts in cognitive science, were already familiar to the participants at the Hixon Symposium and to other scholars involved in the first concerted efforts to found cognitive science during the 1940s and 1950s. Now it was time to put these ideas to optimal scientific use.

Key Theoretical Inputs to Cognitive Science

Mathematics and Computation

The years around the turn of the century were of exceptional importance in mathematics and logic. For nearly two thousand years, the logic of syllogistic reasoning developed in classical times by Aristotle had held sway; but thanks to the work of the German logician Gottlob Frege, a new form of logic, which involved the manipulation of abstract symbols, began to evolve toward the end of the nineteenth century. Then, in the early twentieth century, as I shall elaborate in chapter 4, the British mathematical logicians Bertrand Russell and Alfred North Whitehead sought, with considerable success, to reduce the basic laws of arithmetic to propositions of elementary logic. The Whitehead-Russell work influenced a whole generation of mathematically oriented thinkers, including both Norbert Wiener and John von Neumann, two of the most important contributors to the founding of cognitive science.

In the 1930s, the logical-mathematical work that ultimately had the greatest import for cognitive science was being carried out by a then relatively unknown British mathematician, Alan Turing. In 1936, he developed the notion of a simple machine (subsequently dubbed a "Turing machine") which could in principle carry out any possible conceivable calculation. The notions underlying this "theoretical" machine were simple. All one needed was an infinitely long tape which could pass through the machine and a scanner to read what was on the tape. The tape itself was divided into identical squares, each of which contained upon it either a blank or some kind of slash. The machine could carry out four moves with the tape: move to the right, move to the left, erase the slash, or print the slash. With just these simple operations, the machine could execute any kind of program or plan that could be expressed in a binary code (for example, a code of blanks and slashes). More generally, so long as one could express clearly the steps needed to carry out a task, it could be programmed and carried out by the Turing machine, which would simply scan the tape (no matter what its length) and carry out the instructions (Davis 1958; McCorduck 1979).

Turing's demonstration—and the theorem he proved—was of profound importance for those researchers interested in computing devices. It suggested that a binary code (composed simply of zeros and ones) would make possible the devising and execution of an indefinite number of programs, and that machines operating on this principle could be constructed. As Turing himself pondered computing devices, he became increasingly enthusiastic about their possibilities. In fact, in 1950 (shortly before his untimely death by suicide in his early forties) he suggested that one could so program a machine that it would be impossible to discriminate *its* answers to an interlocutor from those contrived by a living human being—a notion immortalized as the "Turing machine test." This test is used to refute anyone who doubts that a computer can really think: if an observer cannot distinguish the responses of a programmed machine from those of a human being, the machine is said to have passed the Turing test (Turing 1963).

The implications of these ideas were quickly seized upon by scientists interested in human thought, who realized that if they could describe with precision the behavior or thought processes of an organism, they might be able to design a computing machine that operated in identical fashion. It thus might be possible to test on a computer the plausibility of notions about how a human being actually functions, and perhaps even to construct machines about which one could confidently assert that they think just like human beings.

In building upon Turing's ideas, John von Neumann pursued the notion of devising a program to instruct the Turing machine to reproduce itself. Here was the powerful idea of a *stored program:* that is, the computer could be controlled through a program that itself was stored within the computer's internal memory, so that the machine would not have to be laboriously reprogrammed for each new task (see Goldstine 1972). For the first time, it became conceivable that a computer might prepare and execute its own programs.

The Neuronal Model

A second line of thinking important for those involved in founding cognitive science was put forth during the early 1940s by Warren McCulloch, the second speaker at the Hixon Symposium, and Walter Pitts, a young logician. Again, the core idea was disarmingly simple, though the actual mathematical analysis was anything but trivial. McCulloch and Pitts (1943) showed that the operations of a nerve cell and its connections with other nerve cells (a so-called neural network) could be modeled in terms of logic. Nerves could be thought of as logical statements, and the all-or-none property of nerves firing (or not firing) could be compared to the operation of the propositional calculus (where a statement is either true or false) (Heims 1980, p. 211). This model allowed one to think of a neuron as being activated and then firing another neuron, in the same way that an element or a proposition in a logical sequence can imply some other proposition: thus, whether one is dealing with logic or neurons, entity A plus entity B can imply entity C. Moreover, the analogy between neurons and logic could be thought of in electrical terms—as signals that either pass, or fail to pass, through a circuit. The end result of the McCulloch-Pitts demonstration: "Anything that can be exhaustively and unambiguously described . . . is . . . realizable by a suitable finite neural network" (von Neumann, quoted in Bernstein 1982, p. 68).

The designers of the new computational devices were intrigued by the ideas put forth by McCulloch and Pitts. Thanks to their demonstration, the notion of a Turing machine now looked in two directions—toward a nervous system, composed of innumerable all-or-none neurons; and toward a computer that could realize any process that can be unambiguously described. Turing had demonstrated the possibility in principle of computing machines of great power, while McCulloch and Pitts had demonstrated that at least one redoubtable machine—the human brain—could be thought of as operating via the principles of logic and, thus, as a powerful computer.

Ultimately, McCulloch may have carried his own chain of reasoning too far. He was convinced that fundamental problems of epistemology could be stated and solved only in light of the knowledge of the central nervous system (McCorduck 1979), and he tied his claims about thinking very closely to what was known during his own time about the nervous system. Some commentators even feel that the search by McCulloch and his associates for a direct mapping between logic machines and the nervous system was a regressive element in the development of cognitive science: rather than trying to build machines that mimic the brain at a physiological level, analogies should have been propounded and pursued on a much higher level—for example, between the *thinking* that goes on in human problem solving and the *strategies* embodied in a computer program (McCarthy 1984). On the other hand, it was due in part to McCulloch's own analysis that some of the most important aspects of the nervous system came to be better understood: for he sponsored research on the highly specific properties of individual nerve cells. Moreover, very recently computer scientists have once again been drawing directly on ideas about the nature of and connections among nerve cells (see chapter 10, pp. 318–22). On balance, his polymathic spirit seems to have been a benign catalyst for the growth of cognitive science.

The Cybernetic Synthesis

Even as John von Neumann, working at Princeton, was trying to piece together evidence from mathematics, logic, and the nervous system, mathematician Norbert Wiener was engaged in parallel pursuits at the Massachusetts Institute of Technology (see Heims 1980; Wiener 1964). Even more than von Neumann, Wiener had been a mathematical prodigy and, like his counterpart, had made fundamental discoveries in mathematics while still in his early twenties (Wiener had worked on Brownian motion; von Neumann, on quantum theory). Clearly, in these early choices, both men exhibited a practical bent in their mathematics: further, they aspired to influence the growth of science and technology within their society.

During the 1930s and 1940s, Norbert Wiener, by then ensconced at M.I.T., became involved in a variety of worldly projects. In working on servomechanisms—devices that kept anti-aircraft artillery, guided missiles, and airplanes on course—he had come to think about the nature of feedback and of self-correcting and self-regulating systems, be they mechanical or human. He collaborated closely with Vannevar Bush, who had pioneered in the development of analog computers. Wiener was also struck by the importance of the work of his sometime colleagues McCulloch and Pitts, particularly by the suggestive analogies between a system of logical connections and the human nervous system.

Wiener went beyond all of his contemporaries in his missionary conviction that these various scientific and technological developments cohered. Indeed, in his mind they constituted a new science—one founded on the issues of control and communication, which he deemed to be central in the middle of the twentieth century. He first publicly formulated this point of view in a 1943 paper, "Behavior, Purpose, and Teleology" (Rosenblueth, Wiener, and Bigelow 1943), in which he and his fellow authors put forth the notion that problems of control engineering and communication engineering are inseparable; moreover, they center not on the techniques of electrical engineering, but rather on the much more fundamental notion of the message—"whether this should be transmitted by electrical, mechanical, or nervous means." The authors introduced a then-radical notion: that it is legitimate to speak of machines that exhibit feedback as "striving toward goals," as calculating the difference between their goals and their actual performance, and as then working to reduce those differences. Machines were purposeful. The authors also developed a novel notion of the central nervous system. As Wiener later put it:

> The central nervous system no longer appears as a self-contained organ, receiving inputs from the senses and discharging into the muscles. On the contrary, some of its most characteristic activities are explicable only as circular processes, emerging from the nervous system into the muscles, and re-entering the nervous system through the sense organs, whether they be proprioceptors or organs of the special senses. This seemed to us to mark a new step in the study of that part of neurophysiology which concerns not solely the elementary processes of nerves and synapses but the performance of the nervous system as an integrated whole. (Wiener 1961, p. 8)

The parallels to Lashley's ideas about neural organization—and the challenge to behaviorist reflexology—are striking indeed.

Before long, Wiener had contrived a synthesis of the various interlocking ideas and presented it in the landmark volume *Cybernetics* (first published in 1948, the same year as the Hixon Symposium). He introduced his neologistic science as follows: "We have decided to call the entire field of control and communication theory, whether in the machine or in the animal, by the name Cybernetics" (1961, p. 11). In the following pages, he set down an integrated vision—a linkage of developments in understanding the human nervous system, the electronic computer, and the operation of other machines. And he underscored his belief—echoing von Neumann and McCulloch and Pitts—that the functioning of the living organism and the operation of the new communication machines exhibited crucial parallels. Though Wiener's synthesis was not ultimately the one embraced by cognitive science (it came closer to achieving that exalted status in the Soviet Union), it stands as a pioneering example of the viability of such an interdisciplinary undertaking.

Information Theory

Another key progenitor of cognitive science was Claude Shannon, an electrical engineer at M.I.T. who is usually credited with devising information theory. Already as a graduate student at M.I.T. in the late 1930s, Shannon had arrived at a seminal insight. He saw that the principles of logic (in terms of true and false propositions) can be used to describe the two states (on and off) of electromechanical relay switches. In his master's thesis, Shannon provided an early suggestion that electrical circuits (of the kind in a computer) could embody fundamental operations of thought. I shall describe this work—so crucial for all subsequent work with computers—further in chapter 6.

During the next ten years, working in part with Warren Weaver, Shannon went on to develop the key notion of information theory: that information can be thought of in a way entirely divorced from specific content or subject matter as simply a single decision between two equally plausible alternatives. The basic unit of information is the *bit* (short for "binary digit"): that is, the amount of information required to select one message from two equally probable alternatives. Thus, the choice of a message from among eight equally probable alternatives required three bits of information: the first bit narrowed the choice from one of eight to one of four; the second, from one of four to one of two; the third selects one of the remaining alternatives. Wiener explained the importance of this way of conceptualization: "Information is information, not matter or energy. No materialism which does not admit this can survive at the present day" (Wiener 1961, p. 132).

Thanks to Wiener's insights, it became possible to think of information apart from a particular transmission device: one could focus instead on the efficacy of *any* communication of messages via *any* mechanism, and one could consider cognitive processes apart from any particular embodiment—an opportunity upon which psychologists would soon seize as they sought to describe the mechanisms underlying the processing of any kind of information. Only very recently have cognitive scientists begun to wonder whether they can, in fact, afford to treat all information equivalently and to ignore issues of content.

Neuropsychological Syndromes

A comparable contribution to an incipient cognitive science came from a remote and unexpected scientific corner—the profiles of cognitive incapacities following damage to the human brain. Paradoxically, this area of science relies heavily on the travesties of war. As in the era of the First World War, much was learned during the Second World War about aphasia (language deficit), agnosia (difficulty in recognition), and other forms of mental pathology consequent upon injury to the brain. Laboratories in New York, Oxford, Paris, Berlin, and Moscow were all busily engaged in working with victims of brain damage. When the neuropsychological researchers began to communicate their findings to one another, considerable convergence was noted even across cultural and linguistic boundaries. For instance, aphasia assumed similar forms despite wide differences across languages. There was, it seemed, much more regularity in the organization of cognitive capacities in the nervous system than was allowed for by wholly environmental accounts of mental processes. Furthermore, the patterns of breakdown could not be readily explained in terms of simple stimulus-response disruption. Rather, in many cases, the hierarchy of behavioral responses was altered. For example, in certain forms of aphasia, the general sentence frame was preserved, but subjects could not correctly slot individual words into the frame. In other aphasias, the sentence frame broke down, but individual content words carried meaning. Thus was struck yet another blow against reflex-arc models of thought. At the same time, the particular profiles of abilities and disabilities that emerge in the wake of brain damage provided many pregnant suggestions about how the human mind might be organized in normal individuals.

By the late 1940s, in areas as diverse as communication engineering and neuropsychology, certain themes were emerging principally in the United States, Great Britain, and the Soviet Union. Though I have stressed the American version of this story, comparable accounts could be presented from other national perspectives as well. Scholars in these fields were not only writing but were eagerly meeting with one another to discuss the many exciting new perspectives. Herbert Simon, ultimately one of the founders of cognitive science but then a graduate student at the University of Chicago, recalls a kind of "invisible college" in the 1940s (Simon 1982). He knew McCulloch at Chicago; he knew of Shannon's master's thesis at M.I.T.; he knew that Wiener and von Neumann were working on issues in symbolic logic which had grown out of the philosophical writings of Whitehead, Russell, and Frege. Simon himself was studying at Chicago with Rudolf Carnap, who was then putting forth key notions about the syntax of logic. Such leading biologists (and Hixon symposiasts) as Ralph Gerard, Heinrich Klüver, Roger Sperry, and Paul Weiss were working in nearby laboratories on issues of the nervous system. Many of the same influences were rubbing off during this period on Jerome Bruner, Noam Chomsky, John McCarthy, George Miller, Allen Newell, and other founders of cognitive science.

Catalytic Encounters and Influential Writings

By the 1940s, then, the principal intellectual capital on which cognitive science was to be constructed had already emerged. A few scholars like Norbert Wiener attempted a tentative intellectual synthesis, and more than a few—ranging from students like Herbert Simon to masters like John von Neumann—sensed the imminent emergence of a new field (or fields) of study. There was still the resistance implicit in the behaviorist credo, as well as some doubts that the human mind would be able to study itself as effectively as it had studied matter and genetics; but these factors did not suffice to dampen the enthusiasm of those who sensed the vastness of the prize awaiting the Newton of human cognition.

The intellectual history of this era reveals many meetings among those interested in matters of cognition as well as a significant number of publications that helped to promote a new interdisciplinary science of the mind. It is possible, of course, that cognitive science could have come into being —and perhaps even have assumed its present form—in the absence of these conferences, books, and articles. But particularly when scholars seek to join forces across often remote disciplines, it is crucial for them to have the opportunity to get together regularly, to question one another, and to discover those aspects of scientific method, prejudice, and hunch that are often invisible in the written record.

The Hixon Symposium, then, was but one of many conferences held among cognitively oriented scientists during the 1940s and 1950s. To be sure, it was especially important for our story because of two factors: its linking of the brain and the computer and its relentless challenging of the then-prevalent behaviorism. Nonetheless, in any history of this new field, it is necessary to cite a few other circumstances under which aspiring cognitive scientists met one another.

In the scientific annals of this period, the name of the Josiah P. Macy

Foundation looms large. In the winter of 1944, John von Neumann and Norbert Wiener convened a meeting at Princeton of all those interested in what later came to be called "cybernetics." Present at the Macy-sponsored event were many of the scholars already introduced in this narrative. Wiener later recalled, "At the end of the meeting it had become clear to all that there was a substantial common basis of ideas between the workers in the different fields, that people in each group could already use notions which had been better developed by the others, and that some attempt should be made to achieve a common vocabulary" (1961, p. 15).

Building on these initial contacts, Warren McCulloch arranged with the Macy Foundation in the spring of 1946 for a series of meetings on the problems of feedback. "The idea has been to get together a group of modest size, not exceeding some twenty in number, of workers in various related fields and to hold them together for two successive days in all-day series of informal papers, discussions, and meals together, until they had had the opportunity to thresh out their differences and to make progress in thinking along the same lines" (Wiener 1961, p. 18). Ultimately there were ten such meetings, about one a year, of what was originally the Conference for Circular Causal and Feedback Mechanisms in Biological and Social Systems—soon (and happily) shortened, at Wiener's urging, to the Conference on Cybernetics. In the transcripts of these conferences, one discerns ample evidence of scholars informing one another as well as first intimations of interesting and sometimes unexpected projects. For example, it was in discussions at the Macy meetings that the anthropologist Gregory Bateson first encountered ideas about feedback which he was to mine in his "double-bind" theory of schizophrenia.

Activity was especially intense in the Boston and Princeton areas and in California. During the early 1950s, J. Robert Oppenheimer, director of the Princeton Institute for Advanced Study (of which von Neumann was a permanent member) became interested in the application of some of these new ideas in the field of psychology. He regularly invited a group of psychologists to visit the institute and report on recent developments in their field. Among those who spent a year there were George Miller and Jerome Bruner, gifted young psychologists who would shortly play a fundamental role in the launching of cognitive science.

Again, there was difficult-to-anticipate but promising cross-fertilization of ideas. Oppenheimer was particularly interested in analogies between the problems of perception, as they are viewed by the psychologist, and issues of observation, which had come to loom large in atomic and subatomic physics, once one began to work at the atomic and the subatomic levels. He had been pondering the disturbing implications of the *indeterminacy principle,* according to which it is impossible to ascertain the position and the velocity of a particle without affecting it during the course of measurement. Meanwhile, Bruner had been studying the effects of an observer's attitude and expectations on putatively "objective data." One day Oppenheimer remarked to him, "Perception as you psychologists study it can't, after all, be different from observation in physics, can it?" (quoted in Bruner 1983, pp. 95–96).

In Boston, discussion of these cognitive themes was continuing at M.I.T., and at the associated Lincoln Laboratories, where a group of young engineers and psychologists had assembled to work on applied problems, such as early warning signals in the case of bomb attacks. At nearby Harvard in the prestigious Society of Fellows, the influence of behaviorist thinking was dominant among the senior fellows, but the young junior fellows, including the linguist Noam Chomsky and the mathematician Marvin Minsky, were already proceeding in different (and anti-behaviorist) theoretical directions (Miller 1982). The Ford Foundation, having decided to help stimulate work in the behavioral sciences, established a Center for Advanced Study in the Behavioral Sciences in Palo Alto and also provided funding for a significant proportion (perhaps one third) of all the research psychologists in America. At the Rand Corporation in Southern California, groups of mathematicians and engineers were working on the development of computing machines. Two young scientists, Allen Newell and Herbert Simon, had begun to talk about the possibilities of creating machines that could genuinely think. And, again, there was a British version as well—the Ratio Club, which commenced in 1949. Central to the Ratio Club was the notion of processing information in animals and machines. Members included physiologists, engineers, physicians, and psychologists with interests in the mind or "minding." Turing occasionally attended meetings. The group (which met for several years) had the intriguing rule that any member who reached the rank of full professor must be expelled, because he would then have potential control over other members (McCorduck 1979, p. 78).

In addition to these many face-to-face encounters among those concerned with cognitive matters, there appeared in the late 1940s and early 1950s several books from different quarters which helped to bring the emerging interdisciplinary ideas to wider attention. One such book, perhaps the closest analogy in writing to the Hixon Symposium, was W. Ross Ashby's *Design for a Brain* (1952).

Ashby, a British physician and mathematician, wished to account for human mental activity in a mechanistic manner. He sought to show how, using only logical axiomatic methods, one could design a machine capable of adaptive behavior or learning. In the proper behaviorist fashion of the day, Ashby deliberately avoided talking of anything like consciousness or purposeful behavior. Instead, he directed his attention to the way in which an organism can effect a transition from chaos to stability, thereby enhancing the possibility of survival. Stability can come about because "the machine is a self-organizing system, a system that responds to stimuli, changing its behavior and sometimes its shape in order to achieve stability" (McCorduck 1979, p. 83). Ashby's work intrigued young scholars—like George Miller, Marvin Minsky, Allen Newell, and Herbert Simon—for he was not interested merely in making a machine that worked well. "My aim," Ashby declared, "is simply to copy the living brain. In particular if the living brain fails in certain characteristic ways, then I want my artificial brain to fail too. I have attempted to deduce what is necessary, what properties the nervous system must have if it is to behave at once mechanistically and adaptively" (1952, pp. v, 130). It was the scope of Ashby's aspirations, the doggedly logical way in which he proceeded, and his refusal to "finesse" possible differences between human and mechanical behavior which caught the attention of aspiring cognitive scientists. Indeed, Ashby's maddening adherence to the strictest behaviorist and mechanistic canons served as an additional spur to younger investigators: his challenge continues to hang, at least spiritually, above the desks of many of today's cognitive scientists.

From more remote quarters began to appear books relevant to the discussions in the emerging cognitive sciences. For instance, in the area of linguistics, Roman Jakobson and his colleagues published their first findings about the distinctive features of language—the units or building blocks out of which the phonemes (or basic sounds) of language are constructed (Jakobson and Halle 1956). In neuropsychology, Donald Hebb described the developing nervous system so as to account for many aspects of visual perception and also to illuminate processes of learning and the growth and subsequent decline of intelligence (Hebb 1949). In anthropology, Gregory Bateson introduced his notions about feedback systems in social systems—for example, among members of a family (Bateson et al. 1956). New mathematical innovations, such as Markov processes and stochastic models, quickly came to the attention of young workers in the social sciences. And a few names which had garnered attention on the Continent began to command increasing respect in the Anglo-American community—Frederic Bartlett, Claude Lévi-Strauss, Alexander Luria, Jean Piaget, Lev Vygotsky.

But all this is by way of stagesetting. The basic ideas for cognitive science were immanent in the early papers of McCulloch, Turing, von Neumann, Wiener, and Pitts and were being heatedly debated at the Macy conferences, the Ratio Club, Harvard's Society of Fellows, and various other institutions and venues. Important papers and books were being written and discussed. Still, all of this activity was going on, in a sense, outside established fields of study. It was extracurricular and considered a bit odd by those in the mainstream—behaviorist psychology, structural linguistics, functionalist social anthropology, the neuropsychology of animal learning. It would take more dramatic events to shake these fields to their foundation—events that were not long in coming.

3

Cognitive Science: The First Decades

A Consensual Birthdate

Seldom have amateur historians achieved such consensus. There has been nearly unanimous agreement among the surviving principals that cognitive science was officially recognized around 1956. The psychologist George A. Miller (1979) has even fixed the date, 11 September 1956.

Why this date? Miller focuses on the Symposium on Information Theory held at the Massachusetts Institute of Technology on 10–12 September 1956 and attended by many leading figures in the communication and the human sciences. The second day stands out in Miller's mind because of two featured papers. The first, presented by Allen Newell and Herbert Simon, described the "Logic Theory Machine," the first complete proof of a theorem ever carried out on a computing machine. The second paper, by the young linguist Noam Chomsky, outlined "Three Models of Language." Chomsky showed that a model of language production derived from Claude Shannon's information-theoretical approach could not possibly be applied successfully to "natural language," and went on to exhibit his own approach to grammar, based on linguistic transformations. As Miller recalls, "Other linguists had said language has all the formal precisions of mathematics, but Chomsky was the first linguist to make good on the claim. I think that was what excited all of us" (1979, p. 8). Not incidentally, that day George Miller also delivered a seminal paper, outlining his claim that the capacity of human short-term memory is limited to approximately seven entries. Miller summed up his reactions:

I went away from the Symposium with a strong conviction, more intuitive than rational, that human experimental psychology, theoretical linguistics, and computer simulation of cognitive processes were all pieces of a larger whole, and that the future would see progressive elaboration and coordination of their shared concerns. . . . I have been working toward a cognitive science for about twenty years beginning before I knew what to call it. (1979, p. 9)

Miller's testimony is corroborated by other witnesses. From the ranks of psychology, Jerome Bruner declares, "New metaphors were coming into being in those mid-1950s and one of the most compelling was that of computing. . . . My "Generation" created and nurtured the Cognitive Revolution—a revolution whose limits we still cannot fathom" (1983, pp. 274, 277). Michael Posner concludes, "This mix of ideas about cognition was ignited by the information processing language that arrived in psychology in the early 1950s" (Posner and Shulman 1979, p. 374). And George Mandler suggests:

For reasons that are obscure at present, the various tensions and inadequacies of the first half of the twentieth century cooperated to produce a new movement in psychology that first adopted the label of information processing and after became known as modern cognitive psychology. And it all happened in the five year period between 1955 and 1960. Cognitive science started during that five year period, a happening that is just beginning to become obvious to its practitioners. (1981, p. 9)

Finally, in their history of the period, computer scientists Allen Newell and Herbert Simon declare:

Within the last dozen years a general change in scientific outlook has occurred, consonant with the point of view represented here. One can date the change roughly from 1956: in psychology, by the appearance of Bruner, Goodnow, and Austin's *Study of Thinking* and George Miller's "The magical number seven"; in linguistics, by Noam Chomsky's "Three models of language"; and in computer science, by our own paper on the Logical Theory Machine. (1972, p. 4)

This impressive congruence stresses a few seminal publications, emanating (not surprisingly perhaps) from the same small group of investigators. In fact, however, the list of relevant publications is almost endless. As far as general cognitive scientific publications are concerned, John von Neumann's posthumous book, *The Computer and the Brain* (1958), should head the list. In this book—actually a set of commissioned lectures which von Neumann became too ill to deliver—the pioneering computer scientist developed many of the themes originally touched upon in his Hixon Symposium contribution. He included a discussion of various kinds of computers and analyzed the idea of a program, the operation of memory in computers, and the possibility of machines that replicate themselves.

Relevant research emanated from each of the fields that I have designated as contributing cognitive sciences.* The witnesses I have just quoted noted the principal texts in the fields of psychology, linguistics, and artificial intelligence, and many more entries could be added. Neuroscientists were beginning to record impulses from single neurons in the nervous system. At M.I.T., Warren McCulloch's research team, led by the neurophysiologists Jerome Lettvin and Humberto Maturana, recorded from the retina of the frog. They were able to show that neurons were responsive to extremely specific forms of information such as "bug-like" small dark spots which moved across their receptive fields, three to five degrees in extent. Also in the late 1950s, a rival team of investigators, David Hubel and Torsten Wiesel at Harvard, began to record from cells in the visual cortex of the cat. They located nerve cells that responded to specific information, including brightness, contrast, binocularity, and the orientation of lines. These lines of research, eventually honored in 1981 by a Nobel Prize, called attention to the extreme specificity encoded in the nervous system.

The mid 1950s were also special in the field of anthropology. At this time, the first publications by Harold Conklin, Ward Goodenough, and Floyd Lounsbury appeared in the newly emerging field of cognitive anthropology, or ethnosemantics. Researchers undertook systematic collection of data concerning the naming, classifying, and concept-forming abilities of people living in remote cultures, and then sought to describe in formal terms the nature of these linguistic and cognitive practices. These studies documented the great variety of cognitive practices found around the world, even as they strongly suggested that the relevant cognitive processes are similar everywhere.

In addition, in the summer of 1956, a group of young scholars, trained in mathematics and logic and interested in the problem-solving potentials of computers, gathered at Dartmouth College to discuss their mutual interests. Present at Dartmouth were most of the scholars working in what came to be termed "artificial intelligence," including the four men generally deemed to be its founding fathers: John McCarthy, Marvin Minsky, Allen Newell, and Herbert Simon. During the summer institute, these scientists, along with other leading investigators, reviewed ideas for programs that would solve problems, recognize patterns, play games, and reason logically, and laid out the principal issues to be discussed in coming years. Though no synthesis emerged from these discussions, the participants seem to have set up a permanent kind of "in group" centered at the

*Full bibliographical references to these lines of research will be provided at appropriate points in the text.

M.I.T., Stanford, and Carnegie-Mellon campuses. To artificial intelligence, this session in the summer of 1956 was as central as the meeting at M.I.T. among communication scientists a few months later.

Scholars removed from empirical science were also pondering the implications of the new machines. Working at Princeton, the American philosopher Hilary Putnam (1960) put forth an innovative set of notions. As he described it, the development of Turing-machine notions and the invention of the computer helped to solve—or to dissolve—the classical mind-body problem. It was apparent that different programs, on the same or on different computers, could carry out structurally identical problem-solving operations. Thus, the logical operations themselves (the "software") could be described quite apart from the particular "hardware" on which they happened to be implemented. Put more technically, the "logical description" of a Turing machine includes no specification of its physical embodiment.

The analogy to the human system and to human thought processes was clear. The human brain (or "bodily states") corresponded to the computational hardware; patterns of thinking or problem solving ("mental states") could be described entirely separately from the particular constitution of the human nervous system. Moreover, human beings, no less than computers, harbored programs; and the same symbolic language could be invoked to describe programs in both entities. Such notions not only clarified the epistemological implications of the various demonstrations in artificial intelligence; they also brought contemporary philosophy and empirical work in the cognitive sciences into much closer contact.

One other significant line of work, falling outside cognitive science as usually defined, is the ethological approach to animal behavior which had evolved in Europe during the 1930s and 1940s thanks to the efforts of Konrad Lorenz (1935) and Niko Tinbergen (1951). At the very time that American comparative psychologists were adhering closely to controlled laboratory settings, European ethologists had concluded that animals should be studied in their natural habitat. Observing carefully under these naturalistic conditions, and gradually performing informal experiments on the spot, the ethologists revealed the extraordinary fit between animals and their natural environment, the characteristic *Umwelt* (or world view) of each species, and the particular stimuli (or releasers) that catalyze dramatic developmental milestones during "critical" or "sensitive" periods. Ethology has remained to some extent a European rather than an American specialty. Still, the willingness to sample wider swaths of behavior in naturally occurring settings had a liberating influence on the types of concept and the modes of exploration that came to be tolerated in cognitive studies.

The 1960s: Picking Up Steam

The seeds planted in the 1950s sprouted swiftly in the next decade. Governmental and private sources provided significant financial support. Setting the intellectual tone were the leading researchers who had launched the key lines of study of the 1950s, as well as a set of gifted students who were drawn to the cognitive fields, much in the way that physics and biology had lured the keenest minds of earlier generations. Two principal figures in this "selling of cognition" were Jerome Bruner and George Miller, who in 1960 founded at Harvard the Center for Cognitive Studies. The Center, as story has it, began when these two psychologists approached the dean of the faculty, McGeorge Bundy, and asked him to help create a research center devoted to the nature of knowledge. Bundy reportedly responded, "And how does that differ from what Harvard University does?" (quoted in Bruner 1983, p. 123). Bundy gave his approval, and Bruner and Miller succeeded in getting funds from the Carnegie Corporation, whose president at that time, the psychologist John Gardner, was sympathetic to new initiatives in the behavioral sciences.

Thereafter, for over ten years, the Harvard Center served as a locale where visiting scholars were invited for a sabbatical, and where graduate and postdoctorate students flocked in order to sample the newest thinking in the cognitive areas. A list of visitors to the Center reads like a Who's Who in Cognitive Science: nearly everyone visited at one time or another, and many spent a semester or a year in residence. And while the actual projects and products of the Center were probably not indispensable for the life of the field, there is hardly a younger person in the field who was not influenced by the Center's presence, by the ideas that were bandied about there, and by the way in which they were implemented in subsequent research. Indeed, psychologists Michael Posner and Gordon Shulman (1979) locate the inception of the cognitive sciences at the Harvard Center.

During the 1960s, books and other publications made available the ideas from the Center and from other research sites. George Miller—together with his colleagues Karl Pribram, a neuroscientist, and Eugene Galanter, a mathematically oriented psychologist—opened the decade with a book that had a tremendous impact on psychology and allied fields—a slim volume entitled *Plans and the Structure of Behavior* (1960). In it the authors sounded the death knell for standard behaviorism with its discredited reflex arc and, instead, called for a cybernetic approach to behavior in terms of actions, feedback loops, and readjustments of action in the light of feedback. To replace the reflex arc, they proposed a unit of activity called a "TOTE unit" (for "Test-Operate-Test-Exit"): an important property of a TOTE unit was that it could itself be embedded within the hierarchical structure of an encompassing TOTE unit. As a vehicle for conceptualizing such TOTE units, the authors selected the computer with its programs. If a computer could have a goal (or a set of goals), a means for carrying out the goal, a means for verifying that the goal has been carried out, and then the option of either progressing to a new goal or terminating behavior, models of human beings deserved no less. The computer made it legitimate in theory to describe human beings in terms of plans (hierarchically organized processes), images (the total available knowledge of the world), goals, and other mentalistic conceptions; and by their ringing endorsement, these three leading scientists now made it legitimate in practice to abandon constricted talk of stimulus and response in favor of more open-ended, interactive, and purposeful models.

The impact of this way of thinking became evident a few years later when textbooks in cognitive psychology began to appear. By far the most influential was *Cognitive Psychology* by the computer-literate experimental psychologist Ulric Neisser (1967). Neisser put forth a highly "constructive" view of human activity. On his account, all cognition, from the first moment of perception onward, involves inventive analytic and synthesizing processes. He paid tribute to computer scientists for countenancing talk of an "executive" and to information scientists for discussing accession, processing, and transformation of data. But at the same time, Neisser resisted uncritical acceptance of the computer-information form of analysis. In his view, objective calculation of how many bits of information can be processed is not relevant to psychology, because human beings are selective in their attention as a pure channel such as a telephone cannot be. Neisser expressed similar skeptical reservations about the claims surrounding computer programs:

> None of [these programs] does even remote justice to the complexity of human mental processes. Unlike men, "artificially intelligent" programs tend to be single minded, undistractable, and unemotional. . . . This book can be construed as an extensive argument against models of this kind, and also against other simplistic theories of the cognitive processes. (1967, p. 9)

After Neisser, it was possible to buy the cognitive science approach in general and still join into vigorous controversies with "true believers."

Enthusiasts of the power of simulation were scarcely silent during this period. In his 1969 Compton lectures, *The Sciences of the Artificial,* Herbert Simon provided a philosophical exposition of his approach: as he phrased it, both the computer and the human mind should be thought of as "symbol systems"—physical entities that process, transform, elaborate, and, in other ways, manipulate symbols of various sorts. And, in 1972, Allen Newell and Herbert Simon published their magnum opus, the monumental *Human Problem Solving,* in which they described the "general problem solver" programs, provided an explanation of their approach to cognitive studies, and included a historical addendum detailing their claims to primacy in this area of study.

Textbooks and books of readings were appearing in other subfields of cognitive science as well. An extremely influential collection was Jerry Fodor and Jerrold Katz's collection, *The Structure of Language* (1964), which anthologized articles representing the Chomskian point of view in philosophy, psychology, and linguistics, and attempted to document why this approach, rather than earlier forays into language, was likely to be the appropriate scientific stance. In artificial intelligence, Edward Feigenbaum and Julian Feldman put out a collection called *Computers and Thought* (1963), which presented many of the best-running programs of the era; while their collection had a definite "Carnegie slant," a rival anthology, *Semantic Information Processing,* edited by Marvin Minsky in 1968, emphasized the M.I.T. position. And, in the area of cognitive anthropology, in addition to influential writings by Kimball Romney and Roy D'Andrade (1964), Stephen Tyler's textbook *Cognitive Anthropology* made its debut in 1969.

But by 1969, the number of slots in short-term memory had been exceeded—without the benefit of chunking, one could no longer enumerate the important monographs, papers, and personalities in the cognitive sciences. (In fact, though my list of citations may seem distressingly long, I have really only scratched the surface of cognitive science, circa 1970.) There was tremendous activity in several fields, and a feeling of definite progress as well. As one enthusiastic participant at a conference declared:

> We may be at the start of a major intellectual adventure: somewhere comparable to the position in which physics stood toward the end of the Renaissance, with lots of discoveries waiting to be made and the beginning of an inkling of an idea of how to go about making them. It turned out, in the case of the early development of modern physics that the advancement of the science involved developing new kinds of intellectual sophistication: new mathematics, a new ontology, and a new view of scientific method. My guess is that the same sort of evolution is required in the present case (and, by the way, in much the same time scale). Probably now, as then, it will be an uphill battle against obsolescent intellectual and institutional habits. (Sloan Foundation 1976, p. 10)

When the amount of activity in a field has risen to this point, with an aura of excitement about impending breakthroughs, human beings often found some sort of an organization or otherwise mark the new enterprise. Such was happening in cognitive science in the early and middle 1970s. The moment was ripe for the coalescing of individuals, interests, and disciplines into an organizational structure.

The Sloan Initiative

At this time, fate intervened in the guise of a large New York–based private foundation interested in science—the Alfred P. Sloan Foundation. The Sloan Foundation funds what it terms "Particular Programs," in which it invests a sizable amount of money in an area over a few years' time, in the hope of stimulating significant progress. In the early 1970s, a Particular Program had been launched in the neurosciences: a collection of disciplines that explore the nervous system—ranging from neuropsychology and neurophysiology to neuroanatomy and neurochemistry. Researchers drawn from disparate fields were stimulated by such funding to explore common concepts and common organizational frameworks. Now Sloan was casting about for an analogous field, preferably in the sciences, in which to invest a comparable sum.

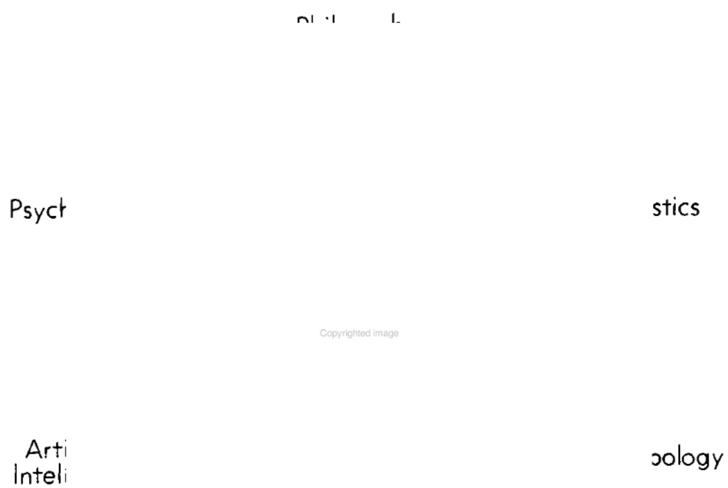
From conversations with officers of the Sloan Foundation, and from the published record, it is possible to reconstruct the principal events that led to the Sloan Foundation's involvement with cognitive science. In early 1975, the foundation was contemplating the support of programs in several fields; but by late 1975, a Particular Program in the cognitive sciences was the major one under active consideration. During the following year, meetings were held where major cognitive scientists shared their views. Possibly sensing the imminent infusion of money into the field, nearly every scientist invited by the Sloan Foundation managed to juggle his or her schedule to attend the meetings. Though there was certainly criticism voiced of the new cognitive science movement, most participants (who were admittedly interested parties) stressed the promise of the field and the need for flexible research and training support.

While recognizing that cognitive science was not as mature as neuroscience at the time of the foundation's commitment to the latter field, officers concluded that "nonetheless, there is every indication, confirmed by the many authorities involved in primary explorations, that many areas of the cognitive sciences are converging, and, moreover, there is a correspondingly important need to develop lines of communication from area to area so that research tools and techniques can be shared in building a body of theoretical knowledge" (Sloan Foundation 1976, p. 6). After deliberating, the foundation decided to embark on a five-to-seven–year program, involving commitments of up to fifteen million dollars. (This commitment was ultimately increased to twenty million dollars.) The investment took the form, initially, of small grants to many research institutions and, ultimately, of a few large-scale grants to major universities.

Like the spur provided by the Macy Foundation a generation earlier, the Sloan Foundation's initiative had a catalytic effect on the field. As more than one person quipped, "Suddenly I woke up and discovered that I had been a cognitive scientist all of my life." In short order the journal *Cognitive Science* was founded—its first issue appearing in January 1977; and soon thereafter, in 1979, a society of the same name was founded. Donald Norman of the University of California in San Diego was instrumental in both endeavors. The society held its first annual meeting, amid great fanfare, in La Jolla, California, in August 1979. Programs, courses, newsletters, and allied scholarly paraphernalia arose around the country and abroad. There were even books about the cognitive sciences, including a popular account, *The Universe Within,* by Morton Hunt (1982) and my own historical essay, also supported by the Sloan Foundation.

Declaring the birth of a field had a bracing effect on those who discovered that they were in it, either centrally or peripherally, but by no means ensured any consensus, let alone appreciable scientific progress. Patrons are almost always necessary, though they do not necessarily suffice, to found a field or create a consensus. Indeed, tensions about what the field is, who understands it, who threatens it, and in what direction it ought to go were encountered at every phase of the Sloan Foundation's involvement (and have continued to be to this day).

Symptomatic of the controversy engendered by the Sloan Foundation's support of research in cognitive science was the reaction to a report commissioned by the foundation in 1978. This State of the Art Report (soon dubbed "SOAP" for short) was drafted by a dozen leading scholars in the field, with input from another score of advisers. In the view of the authors, "What has brought the field into existence is a common research objective: to discover the representational and computational capacities of the mind and their structural and functional representation in the brain" (1978, p. 6). The authors prepared a sketch of the interrelations among the six constituent fields—the cognitive hexagon, as it was labeled. Through the use of unbroken and broken lines, an effort was made to indicate the connections between fields which had already been forged, and to suggest the kinds of connection which could be but had not yet been effected.



Connections among the Cognitive Sciences
KEY: Unbroken lines = strong interdisciplinary ties
Broken lines = weak interdisciplinary ties

In my view, the authors of the SOAP document made a serious effort to survey principal lines of research and to provide a general charter for work in cognitive science, setting forth its principal assumptions. Then, using the example of how individuals from different cultures give names to colors, these authors illustrated how different disciplines combine their insights. (I'll flesh out this example of color naming in chapter 12.) However, the community-at-large adopted a distinctly negative view of the report. In fact, such virulent opposition was expressed by so many readers that, counter to original plans, the document was never published. I think this negative reaction came from the fact that each reader approached the document from the perspective of his or her own discipline and research program. In an effort to be reasonably ecumenical, the authors simply ensured that most readers would find their own work slighted. Moreover, there is as yet no agreed-upon research paradigm—no consensual set of assumptions or methods—and so cognitive scientists tend to project their own favorite paradigms onto the field as a whole. In view of these factors, it was probably not possible in 1978 to write a document that would have won the support of a majority of cognitive scientists.

It would be desirable, of course, for a consensus mysteriously to emerge, thanks to the largesse of the Sloan Foundation, or for some latter-day Newton or Darwin to bring order into the field of cognitive science. In the absence, however, of either of these miraculous events, it is left to those of us who wish to understand cognitive science to come up with our own tentative formulation of the field. In the opening chapter of this book, I presented a working definition of cognitive science and alluded to five key components of the field. Now that I have sketched out some of the intellectual forces that led to the launching of cognitive science some three decades ago, I want to revisit these themes in somewhat more detail, in order to consider some of their implications as well as some of their problematic aspects. I will then conclude this introductory part by describing the paradox and the challenge standing at the center of contemporary cognitive science.

Key Features of Cognitive Science

In my own work I have found it useful to distinguish five features or "symptoms" of cognitive science: the first two of these represent the "core assumptions" of the field, while the latter three represent methodological or strategic features. Not only are these ideas common to most "strong versions" of cognitive science, but they also serve as specific points of contention for its critics. I shall list each of these characteristics and then indicate certain lines of criticism put forth by those most antagonistic to cognitive science. These criticisms (as voiced by their most vocal adherents) will be expanded upon at appropriate points in the book and reviewed in my concluding chapter.

Representations

Cognitive science is predicated on the belief that it is legitimate—in fact, necessary—to posit a separate level of analysis which can be called the "level of representation." When working at this level, a scientist traffics in such representational entities as symbols, rules, images—the stuff of representation which is found between input and output—and in addition, explores the ways in which these representational entities are joined, transformed, or contrasted with one another. This level is necessary in order to explain the variety of human behavior, action, and thought.

In opting for a representational level, the cognitive scientist is claiming that certain traditional ways of accounting for human thought are inadequate. The neuroscientist may choose to talk in terms of nerve cells, the historian or anthropologist in terms of cultural influences, the ordinary person or the writer of fiction in terms of the experiential or phenomenological level. While not questioning the utility of these levels for various purposes, the cognitive scientist rests his discipline on the assumption that, for scientific purposes, human cognitive activity must be described in terms of symbols, schemas, images, ideas, and other forms of mental representation.

In terms of ordinary language, it seems unremarkable to talk of human beings as having ideas, as forming images, as manipulating symbols, images, or languages in the mind. However, there is a huge gap between the use of such concepts in ordinary language and their elevation to the level of acceptable scientific constructs. Cautious theorists want to avoid positing elements or levels of explanation except when absolutely necessary; and they also want to be able to describe the structure and the mechanisms employed at a level before "going public" with its existence. While talk about the structure and mechanisms of the nervous system is relatively unproblematic—since its constituent units can (at least in principle) be seen and probed—agreement to talk of structure and processes at the level of mental representation has proved far more problematic.

Critics of the representational view are generally drawn from behaviorist ranks. Wielders of Ockham's razor, they believe that the construct of mind does more harm than good; that it makes more sense to talk about neurological structures or about overt behaviors, than about ideas, concepts, or rules; and that dwelling on a representational level is unnecessary, misleading, or incoherent.

Another line of criticism, less extreme but ultimately as crippling, accepts the need for common-sense talk about plans, intentions, beliefs, and the like but sees no need for a separate scientific language and level of analysis concerned with their mental representation: on this point of view, one should be able to go directly from plans to the nervous system, because it is *there,* ultimately, that all plans or intentions must be represented. Put in a formula, ordinary language plus neurology eliminate the need for talk of mental representations.

Of course, among scholars who accept the need for a level of representation, debates still rage. Indeed, contemporary theoretical talk among "card-carrying" cognitive scientists amounts, in a sense, to a discussion of the best ways of conceptualizing mental representations. Some investigators favor the view that there is but a single form of mental representation (usually, one that features propositions or statements); some believe in at least two forms of mental representation—one more like a picture (or image), the other closer to propositions; still others believe that it is possible to posit multiple forms of mental representation and that it is impossible to determine which is the correct one.

All cognitive scientists accept the truism that mental processes are ultimately represented in the central nervous system. But there is deep disagreement about the relevance of brain science to current work on cognition. Until recently, the majority viewpoint has held that cognitive science is best pursued apart from detailed knowledge of the nervous system—both because such knowledge has not yet been forthcoming and out of a desire to ensure the legitimacy of a separate level of mental representation. As the cognitive level becomes more secure, and as more discoveries are made in the brain sciences, this self-styled distancing may be reduced. Not surprisingly, neuroscientists (as a group) have shown the least enthusiasm for a representational account, whereas such an account is an article of faith among most psychologists, linguists, and computer scientists.

Computers

While not all cognitive scientists make the computer central to their daily work, nearly all have been strongly influenced by it. The computer serves, in the first place, as an "existence-proof": if a man-made machine can be said to reason, have goals, revise its behavior, transform information, and the like, human beings certainly deserve to be characterized in the same way. There is little doubt that the invention of computers in the 1930s and 1940s, and demonstrations of "thinking" in the computer in the 1950s, were powerfully liberating to scholars concerned with explaining the human mind.

In addition to serving as a model of human thought, the computer also serves as a valuable tool to cognitive scientific work: most cognitive scientists use it to analyze their data, and an increasing number attempt to simulate cognitive processes on it. Indeed, artificial intelligence, the science built around computer simulation, is considered by many the central discipline in cognitive science and the one most likely to crowd out, or render superfluous, other older fields of study.

In principle, it is possible to be a cognitive scientist without loving the computer; but in practice, skepticism about computers generally leads to skepticism about cognitive science. To some critics, computers are just the latest of a long series of inadequate models of human cognition (remember the switchboard, the hydraulic pump, or the hologram) and there is no reason to think that today's "buzz-model" will meet a happier fate. Viewing active organisms as "information-processing systems" seems a radical mistake to such critics. Computers are seen by others as mere playthings which interfere with, rather than speed up, efforts to understand human thought. The fact that one can simulate any behavior in numerous ways may actually impede the search for the correct description of *human* behavior and thought. The excessive claims made by proponents of artificial intelligence are often quoted maliciously by those with little faith in man-made machines and programs.

Involvement with computers, and belief in their relevance as a model of human thought, is pervasive in cognitive science; but again, there are differences across disciplines. Intrinsic involvement with computers is a reliable gauge of the extent of a discipline's involvement with cognitive science. Computers are central in artificial intelligence, and only a few disgruntled computer scientists question the utility of the computer as a model for human cognition. In the fields of linguistics and psychology, one will encounter some reservations about a computational approach; and yet most practitioners of these disciplines do not bother to pick a feud with computerphiles.

When it comes to the remaining cognitive sciences, however, the relationship to the computer becomes increasingly problematic. Many anthropologists and many neuroscientists, irrespective of whether they happen to use computers in their own research, have yet to be convinced that the computer serves as a viable model of those aspects of cognition in which they are interested. Many neuroscientists feel that the brain will provide the answer in its own terms, without the need for an intervening computer model; many anthropologists feel that the key to human thought lies in historical and cultural forces that lie external to the human head and are difficult to conceptualize in computational terms. As for philosophers, their attitudes toward computers range from unabashed enthusiasm to virulent skepticism—which makes them a particularly interesting and important set of informants in any examination of cognitive science.

De-Emphasis on Affect, Context, Culture, and History

Though mainstream cognitive scientists do not necessarily bear any animus against the affective realm, against the context that surrounds any action or thought, or against historical or cultural analyses, in practice they attempt to factor out these elements to the maximum extent possible. So even do anthropologists when wearing their cognitive science hats. This may be a question of practicality: if one were to take into account these individualizing and phenomenalistic elements, cognitive science might become impossible. In an effort to explain everything, one ends up explaining nothing. And so, at least provisionally, most cognitive scientists attempt to so define and investigate problems that an adequate account can be given without resorting to these murky concepts.

Critics of cognitivism have responded in two principal ways. Some critics hold that factors like affect, history, or context will *never* be explicable by science: they are inherently humanistic or aesthetic dimensions, destined to fall within the province of other disciplines or practices. Since these factors are central to human experience, any science that attempts to exclude them is doomed from the start. Other critics agree that some or all of these features are of the essence in human experience, but do not feel that they are insusceptible to scientific explanation. Their quarrel with an antiseptic cognitive science is that it is wrong to bracket these dimensions artificially. Instead, cognitive scientists should from the first put their noses to the grindstone and incorporate such dimensions fully into their models of thought and behavior.

Belief in Interdisciplinary Studies

While there may eventually be a single cognitive science, all agree that it remains far off. Investigators drawn from a given discipline place their faith in productive interactions with practitioners from other disciplines: in the tradition of the Hixon and Macy symposiasts, they hope that, working together, they can achieve more powerful insights than have been obtained from the perspective of a single discipline. As examples, they point to current work in visual perception and in linguistic processing which has come to draw quite naturally on evidence from psychology, neuroscience, and artificial intelligence—so much so that disciplinary lines are beginning to blur.

Skeptics feel that you cannot make progress by compounding disciplines, and that it is more prudent to place each individual disciplinary house in order. Since it is also unclear *which* of the relevant disciplines will ultimately contribute to a cognitive science, and in which way, much valuable time may be wasted in ill-considered collaborations. From their vantage point, it is perfectly all right to have individual cognitive sciences but ill-considered to legislate a single seamless discipline. At most, there should be cooperation among disciplines—and never total fusion.

Rootedness in Classical Philosophical Problems

As already indicated, I consider classical philosophical problems to be a key ingredient in contemporary cognitive science and, in fact, find it difficult to conceive of cognitive science apart from them. The debates of the Greek philosophers, as well as of their successors in the Enlightenment, stand out in many pages of cognitive scientific writing. I do not mean that these traditional questions have necessarily been phrased in the best way, or even that they can be answered, but rather that they serve as a logical point of departure for investigations in cognitive science.

In my discussions with cognitive scientists, however, I have found this precept to be contentious. Nor is it predictable which scientists, or which science, will agree with a philosophically based formulation of the new field. Some cognitive scientists from each discipline readily assent to the importance—indeed, the inevitability—of a philosophical grounding; while others find the whole philosophical enterprise of the past irrelevant to their concerns or even damaging to the cognitive scientific effort. We may well be dealing here with personal views about the utility of reading and debating classical authorities rather than with fundamental methodological aspects of cognitive science. But whatever the reason, cognitive scientists are scarcely of a single mind when it comes to the importance of the *Meno,* of Descartes's *Cogito,* or of Kant's *Critique.*

Precisely because the role of philosophy is controversial in the cognitive sciences, it is useful to explore the earlier history of philosophy. Only such a survey can prove that cognitive scientists—whether or not they are fully aware of it—are engaged in tackling those issues first identified by philosophers many decades or even many centuries ago. Scientists will differ on whether these questions were properly formulated, on whether philosophers made any significant progress in answering them, and on whether philosophers today have any proper role in a scientific enterprise. Indeed, even philosophers are divided on these issues. Still, it is worth reviewing their positions on these issues, for philosophers have, since classical times, taken as their special province the definition of human knowledge. Moreover, they have also pondered the nature and scope of the cognitive-scientific enterprise, and their conclusions merit serious examination.

In my own view, each of these symptoms or features of cognitive science were already discernible in the discussions of the 1940s and were widespread by the middle 1950s. A cognitive-science text will not necessarily exhibit or illustrate each of the symptoms, but few texts will be devoid of most of them. What legitimizes talk of cognitive science is the fact that these features were not in evidence a half-century ago; and to the extent that they once again pass from the scene, the era of cognitive science will be at an end.

Comments on the ultimate fate of cognitive science are most properly left to the conclusion of this study; but as a kind of guidepost to succeeding chapters, it may be useful to anticipate my principal conclusions. In my view, the initial intoxication with cognitive science was based on a shrewd hunch: that human thought would turn out to resemble in significant respects the operations of the computer, and particularly the electronic serial digital computer which was becoming widespread in the middle of the century. It is still too early to say to what extent human thought processes are computational in this sense. Still, if I read the signs right, one of the chief results of the last few decades has been to call into question the extent to which higher human thought processes—those which we might consider most distinctively human—can be adequately approached in terms of this particular computational model.

Which leads to what I have termed the *computational paradox.* Paradoxically, the rigorous application of methods and models drawn from the computational realm has helped scientists to understand the ways in which human beings are not very much like these prototypical computers. This is not to say that no cognitive processes are computerlike—indeed, some very much resemble the computer. Even less is it to contend that cognitive processes cannot be modeled on a computer (after all, *anything* that can be clearly laid out can be so modeled). It is rather to report that the kind of systematic, logical, rational view of human cognition that pervaded the early literature of cognitive science does not adequately describe much of human thought and behavior. Cognitive science can still go on, but the question arises about whether one ought to remain on the lookout for more veridical models of human thought.

Even as cognitive science has spawned a paradox, it has also encountered a challenge. It seems clear from my investigation that mainstream cognitive science comfortably encompasses the disciplines of cognitive psychology, artificial intelligence, and large sections of philosophy and linguistics. But it seems equally clear that other disciplines mark a boundary for cognitive science. Much of neuroscience proceeds at a level of study where issues of representation and of the computer-as-model are not encountered. On the opposite end of the spectrum, much of anthropology has become disaffected with methods drawn from cognitive science, and there is a widespread (and possibly growing) belief that the issues most central to anthropology are better handled from a historical or a cultural or even a literary perspective.

And here inheres the challenge to cognitive science. It is important for cognitive science to establish its own autonomy and to demonstrate terrains in which computational and representational approaches are valid. I believe that cognitive science has already succeeded in this endeavor, though the scope of its enterprise may not be so wide as one would have wished.

If cognitive scientists want to give a complete account of the most central features of cognition, however, they (or other scientists) will have to discover or construct the bridges connecting their discipline to neighboring areas of study—and, specifically, to neuroscience at the lower bound, so to speak, and to cultural studies at the upper. How to do this (or whether it can be done at all) is far from clear at this point: but unless the cognitive aspects of language or perception or problem solving can be joined to the neuroscientific and anthropological aspects, we will be left with a disembodied and incomplete discipline. Put differently, no one challenges the autonomy of biology, chemistry, and physics; but unless a single narrative can be woven from the components of atomic, molecular, and organic knowledge, the full nature of organic and inorganic matter will remain obscure.

All this risks getting ahead of our story, however. We have seen in the preceding pages how different factors present early in the century came together to form the bedrock of a new discipline. Ultimately, I want to take a close look at some of the best work in the discipline, so that I can properly evaluate its current status and its future prospects. To achieve this overview, however, it is necessary to consider how the very framing of questions within cognitive science grows out of philosophical writings of the past. By the same token, it is necessary to understand the particular histories, methods, and problems that have characterized the component cognitive sciences. Ultimately this philosophical and historical background has determined in large measure the nature and scope of current interdisciplinary cognitive-scientific efforts. In part II of this book, I shall take a careful look at the several disciplines whose existence made possible the idea of cognitive science and whose practitioners will determine the success of this enterprise.

PSYCHOLOGY/HISTORY

THE MIND'S NEW SCIENCE

Howard Gardner

WITH A NEW EPILOGUE BY THE AUTHOR:
Cognitive Science After 1984

What is the nature of knowledge? A renowned psychologist provides the first full-scale history of cognitive science.

"A breathtakingly wide-ranging exploration of the current problems and future prospects of the field cast against the history, subject matter, and interconnections of its constituent disciplines. . . . a prodigiously ambitious book."—DANIEL J. KEVLES, *The New Republic*

"Splendid. A thrilling, intimate account of three decades of an exciting and historically unparalleled period in the study of mind." —KARL PRIBRAM, Stanford University

"More than a prodigiously researched, fascinating history, this book explores the foundations of cognitive science and the interconnections among its seemingly disparate disciplines."—STEPHEN M. KOSSLYN, Professor of Psychology, Harvard University

"Superb. What an amazing scholar Gardner is! Even if you have been a participant in what he calls the 'cognitive revolution,' you will never have grasped the whole as it is presented in this wonderful book."—ROGER BROWN, Harvard University

HOWARD GARDNER is affiliated with Harvard University, Boston University School of Medicine, and Boston Veterans Administration Medical Center. His numerous books include *Frames of Mind* (Basic, 1983), *Art, Mind, and Brain* (Basic, 1982), and *Artful Scribbles* (Basic, 1980).

COVER DESIGN BY WILLIAM DAVIS

BASIC BOOKS
A Member of the Perseus Books Group
www.basicbooks.com

ISBN-13: 978-0-465-04635-5
ISBN-10: 0-465-04635-5
90000
9 780465 046355