THE MIND'S NEW SCIENCE
A History of the Cognitive Revolution

HOWARD GARDNER

With a New Epilogue by the Author:
Cognitive Science After 1984


---

Library of Congress Cataloging in Publication Data

Gardner, Howard.

The mind’s new science.

Bibliography: p. 401.

Includes index.


ISBN 0-465-04634-7 (cloth)


---

Epilogue to Paperback Edition Copyright © 1987

by Howard Gardner

Copyright © 1985 by Howard Gardner

Printed in the United States of America

Designed by Vincent Torre

89 90 HC 9 8 7 6 5
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:
Cognitive Science: The First Decades

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 comput-
ers 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
Cognitive Science: The First Decades

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.
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 inade-
quate. 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 possi-
ble 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
Cognitive Science: The First Decades

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 explica-
ble 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
antisepic 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 disci-
plines, 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
Cognitive Science: The First Decades

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.
Conclusion: The Computational Paradox and the Cognitive Challenge

Surveying the scientific landscape at the beginning of the century, a far-sighted observer might have felt justified in announcing the arrival of the mind's new science. After all, building on the philosophical tradition of the Greeks and the Enlightenment, and in the wake of dramatic breakthroughs in physics, chemistry, and biology, the solution to the mystery of human mental processes seemed at hand. Moreover, toward the end of the nineteenth century, a raft of new disciplines concerned particularly with human thought and behavior had been launched. Surely the opportunity to look at individuals in many cultures, in the light of the latest findings about the human nervous system and with the powerful tools of logic and mathematics, should sooner or later yield a bona fide science of the mind.

From a contemporary perspective, it seems evident that at least three conditions had to fall into place before this dream could reach fruition. First of all, it was necessary to demonstrate the inadequacies of the behaviorist approach. Second, the particular limitations of each social science had to be acknowledged. Finally, the advent of the computer was needed to provide the final impetus for a new cognitive science.
In the preceding chapters, I have shown how each of these three conditions came to be met. By 1948, when Karl Lashley gave his famous Hixon Symposium address on the problem of serial order in behavior, it had become apparent to many scientists that the behaviorist approach to human intellective activity was fatally flawed. By the same token, the limits of other schools in the behaviorist orbit—logical positivism, structural linguistics, anthropological functionalism, Pavlovian reflexology—were already becoming apparent. A fresh approach to these issues was sorely needed.

Paralleling the discovery of the limitations of the behaviorist stance was a growing realization that each of the several human and behavioral sciences, practiced alone, harbored distinct and possibly crippling limitations. Whether it was philosophy's ambivalence about the relevance of empirical data to long-standing epistemological issues, or psychology's difficulty in adjusting its experimental approaches to large-scale issues, or anthropology's problems in transcending the single case study, or neuroscience's ambitions for dealing with capacities that defy reduction to the neural level, these various sciences increasingly felt the need for fertilization with neighboring disciplines.

Finally, and perhaps most decisively, there was the coalescence of various mathematical and logical demonstrations (such as those of Shannon, Turing, and von Neumann) with important technological breakthroughs, which culminated around mid-century in the first computers. Once the power of these machines for dealing with symbolic materials had been demonstrated, many researchers became convinced that a science of cognition might be fashioned in the image of the computer. By 1956, psychologists such as George Miller and Jerome Bruner, computer scientists such as Allen Newell and Herbert Simon, and linguists such as Noam Chomsky had carried out work that (in retrospect) was cognitive-scientific in spirit. And thirty years later, building on these pioneering efforts, researchers such as David Marr and Stephen Kosslyn (working at the intersection of perceptual psychology and artificial intelligence), Eleanor Rosch (combining psychological and anthropological concerns), and Philip Johnson-Laird (synthesizing approaches drawn from philosophy, psychology, linguistics, and artificial intelligence) had demonstrated that clear progress could be made in resolving long-standing philosophical and scientific issues. Though work on the perceptual issues is further along than research on classification or on rationality, it seems reasonable to declare in 1985 that cognitive science has come of age.

It is therefore opportune, in the life of the science as well as in the course of this survey, to take stock: to revisit the principal themes of cognitive science in order to clarify what has been accomplished over the
Conclusion: The Computational Paradox and the Cognitive Challenge

past few decades and to discern what remains to be accomplished if cognitive science is to achieve its full potential. This evaluation will entail a consideration of the central concept in cognitive science—that of the representational level—as well as a re-examination of two themes introduced in the opening chapters of this book, the computational paradox and the cognitive challenge.

The Centrality of Mental Representation

To my mind, the major accomplishment of cognitive science has been the clear demonstration of the validity of positing a level of mental representation: a set of constructs that can be invoked for the explanation of cognitive phenomena, ranging from visual perception to story comprehension. Where forty years ago, at the height of the behaviorist era, few scientists dared to speak of schemas, images, rules, transformations, and other mental structures and operations, these representational assumptions and concepts are now taken for granted and permeate the cognitive sciences.

While most researchers (and perhaps most readers) take the representational level for granted, this form of analysis must be situated with reference to competing levels of description and analysis. It has long been acceptable in empirical science to talk of the nervous system and, more generally, of biological systems. These can, after all, be seen and even dissected. While most physical scientists have been unconcerned professionally with cultural and historical matters, it has been acceptable (and uncontroversial) among scholars in the humanities and social sciences to offer explanations in terms of social forces, cultural practices, historical traditions, and the like. How else, after all, to deal with macroscopic social phenomena? The triumph of cognitivism has been to place talk of representation on essentially equal footing with these entrenched modes of discourse—with the neuronal level, on the one hand, and with the socio-cultural level, on the other. Whoever wishes to banish the representational level from scientific discourse would be compelled to explain language, problem solving, classification, and the like strictly in terms of neurological and cultural analysis. The discoveries of the last thirty years make such an alternative most unpalatable.

Making the general case for representation is one thing, making it with precision and power quite another. Any number of vocabularies and conceptual frameworks have been constructed in an effort to characterize the representational level—scripts, schemas, symbols, frames, images, mental models, to name just a few. And any number of terms describe the operations carried out upon these mental entities—transformations, con-
junctions, deletions, reversals, and so on. Cognitive science needs to put its conceptual house in order and to transcend slogans and "buzz" words; the field must agree upon a language for talking about a range of representational phenomena—even if that language turns out to harbor various dialects.

As a start, I would single out two varieties of representation. One form is initially or eventually built into the hardware—be it computer or brain. Such a form must be invoked in order to detail what happens to information but this variety of representation does not involve processes of which the organism is in any way conscious or aware. For example, during the early stages of visual processing described by Marr and his colleagues, the visual system must create symbolic representations of physical information and then operate on these representations. But no organism has any options about these steps, and they are accessible only to a cognitive scientist.

A second variety of representation encompasses those problem-solving and classificatory behaviors that individuals carry out with some flexibility and some degree of explicitness and awareness. In analyzing a sentence or a story, in creating an image or transforming it, one may well become aware of having created some mental representation—or mental model—and then one carries out operations upon that model. Explicit awareness is not necessary here, but it is at least a possibility. Moreover, the individual has the option of changing the mode of representation or the kind of rule that is invoked. This mental activity is appropriately described in terms of representational language but clearly warrants a separate status (or terminology) from the kinds of representations that are automatic and possibly wired in.

It may well be that there are several varieties of representation, or that there exists a continuum from implicit to explicit, or from hard-wired to flexibly programmed. But unless a taxonomy can be agreed upon, discussion of representation will seem ad hoc and unsatisfactory. If representation is indeed the linchpin of cognitive science, it must ultimately be stated as clearly and accepted as widely as quantum theory in physics or the genetic code of the biochemical sciences. Such clarity and consensus seem a long way off.

The Computational Paradox

Strictly speaking, one could have had cognitive science without the computer. After all, computational theory antedated the invention of the computer. And yet as a matter of historical fact, cognitive science was unlikely to have arisen when it did, or taken the form that it has, without
Conclusion: The Computational Paradox and the Cognitive Challenge

the emergence of the computer in our time. Since the first generation of cognitive scientists, the computer has served as the most available and the most appropriate model for thinking about thinking. And for most, it soon became indispensible in their daily empirical and theoretical work. Though the linking of computation and cognitivism turns out to have been a contingent rather than a necessary fact, the fate of cognitive science is closely tied to the fate of the computer.

And this leads to that strange state of affairs I have dubbed the computational paradox. With the vigorous tradition, since the time of the Greeks, of thinking about human thought as an embodiment of mathematical principles, it is hardly surprising that the first generation of cognitivists—reared in the logical positivist tradition—should have embraced a highly rationalistic view of human thought. One of the major results of the first years of cognitive science, however, has been a challenge of that ready assumption.

To be sure, when it comes to elementary and relatively “impenetrable” processes like visual perception or syntactic analysis, an authoritative computational account may some day be given. That is, the kinds of descriptions that are legitimately offered in the terms of a digital von Neumann computer may turn out to be appropriate accounts of these human cognitive processes as well. But as one moves to more complex and belief-tainted processes such as classification of ontological domains or judgments concerning rival courses of action, the computational model becomes less adequate. Human beings apparently do not approach these tasks in a manner that can be characterized as logical or rational or that entail step-by-step symbolic processing. Rather, they employ heuristics, strategies, biases, images, and other vague and approximate approaches. The kinds of symbol-manipulation models invoked by Newell, Simon, and others in the first generation of cognitivists do not seem optimal for describing such central human capacities.

The paradox lies in the fact that these insights came about largely through attempts to use computational models and modeling; only through scrupulous adherence to computational thinking could scientists discover the ways in which humans actually differ from the serial digital computer—the von Neumann computer, the model that dominated the thinking of the first generation of cognitive scientists.

I must again underscore one point. By insisting on the computational paradox, I do not mean to assert that it is impossible to arrive at a computational account of human behavioral and thought patterns in all of their perversity, irrationality, and subjectivity. Such accounts may well be possible and—as has been known since the time of Turing—are certainly possible in principle. Rather, the paradox suggests that the portrait of
human cognition emerging from cognitive science is far removed—at least at the molar level—from the orderly, precise, step-by-step image that dominated the thinking of the founders of the field (and of those who dreamed about it in the more distant past). Human thought emerges as messy, intuitive, subject to subjective representations—not as pure and immaculate calculation. These processes may ultimately be modeled by a computer, but the end result will bear little resemblance to that view of cognition canonically lurking in computationally inspired accounts.

Entailed in the reliance on the computer as a pivotal model of thought is another difficulty which has only recently begun to be recognized. Invocation of the computer leads naturally to a concentration on logical problem solving (à la Newell and Simon) or on orderly, highly rule-governed analysis (à la Chomsky). But evidence from neuropsychological and developmental studies of mental processes has indicated that our concepts of cognition need to be considerably broadened. Processes involved in musical and other artistic activities, and, quite possibly, processes involved in knowing other individuals and in knowing oneself merit the modifier cognitive (Gardner 1983). To the extent that this position is valid, a thoroughgoing cognitive science will need to account for these abilities as well as for more familiar logical mathematical applications of mind. Whatever their relevance for the study of human rationality or problem solving, models derived from the computer are even less likely to be adequate to account for these other uses of mind.

It could be countered that cognitive science ought to be satisfied with modeling logical thought and that these other forms of thought ought to fall by the wayside. Perhaps cognitive science should embrace a classical computational account, even if humans do not much resemble a classical kind of computer. To restrict cognitive science to one form of cognition, however, is to refashion the subject matter to fit the current tools of study. By the same token, to accept an account just because such an account can be given is a scientifically weak move. After all, the purpose of science is not to propose a possible analysis (of which there will always be an infinite number) but rather to come up with the analysis that is most appropriate, parsimonious, and convincing. All cognitive phenomena could, after all, be described in terms of atoms, or in terms of historical factors—and in either event, a representational account would not even be necessary. But now, at the very time when a representational account has been accepted, it is important to try to find the optimal representational account. Representation without computation is one possible outcome for certain regions of cognitive science.

The idea of representation has until this point been closely tied to our current conceptions of computers. But there is no way of determin-
Conclusion: The Computational Paradox and the Cognitive Challenge

ing a priori to what extent the ways currently embraced for describing the representations of computers will prove germane to organisms, be they paramecia or professors. The kinds of representations favored by neo-associationists like Geoffrey Hinton turn out to be radically different (and much sparer) than those countenanced by Jerry Fodor or Zenon Pylyshyn; moreover, it may turn out that neither is adequate or suited for describing an individual who is dreaming, writing a poem, or listening to music. Earlier models of thought—the reflex arc, the hydraulic engine, the telephone switchboard—are now seen to be extremely limited. It is already clear that one kind of computer does not suffice to model all thought. We must face the alternative that humans may be an amalgam of several kinds of computers, or computer models, or may deviate from any kind of computer yet described. Computers will be pivotal in helping us determine how computerlike we are, but the ultimate verdict may be “Not very much.”

Even if computers emerge as viable models for certain facets of human thought, the question arises about the various aspects of human nature that have been bracketed by cognitive scientists. As I noted in chapter 3, nearly all cognitive scientists have conspired to exclude from consideration such nontrivial factors as the role of the surrounding context, the affective aspects of experience, and the effects of cultural and historical factors on human behavior and thought (see D. Norman 1980). Some take the position that this is only a temporary move, until the relatively discrete aspects of cognition have been unraveled; others take the stronger positions that cognitive science should never deal with these aspects or even that a cognitive-scientific account will ultimately render unnecessary any account of these “fuzzier” factors.

Even a brief consideration of each of these “bracketed” topics would require many pages, and since cognitive scientists have themselves steered clear of these issues, there is little work within the disciplinary tradition on which I can draw. My own belief is that, ultimately, cognitive science will have to deal with these factors in one of two ways. Either scientists will propose a cognitive account of affect in which, for example, affective states will be viewed as quantitative values along a dimension, like happiness or cruelty; or researchers will opt for a complex explanatory framework in which the interaction of traditional cognitive factors with affective or cultural factors can somehow be modeled. These will be important but enormously difficult undertakings, for which traditional computational considerations may provide scant help.

Scholars differ widely from one another in their intuitions about the extent to which these other factors may ultimately engulf cognitive factors. From the perspective of a philosopher like Hubert Dreyfus, a linguist
like Roy Harris, a psychologist like Benny Shanon, or an anthropologist like Clifford Geertz, these factors are so important, so constitutive of human experience, that they, rather than cognitive factors, ought to be regarded as primary. Although I am not without sympathy for this perspective, my own view is that there is a heartland of cognition which can be accounted for on its own terms, without necessary reference to (or reliance upon) these other, undoubtedly important elements. It is this heartland that I have attempted to describe in this book, particularly in the last four chapters. The borders of this heartland may determine the limits of cognitive science.

My own doubts about the computer as the guiding model of human thought stem from two principal considerations. As Hilary Putnam (1981) has stressed, the community surrounding a cognizing individual is critical. From those around us, we come to understand which sorts of views are considered acceptable, which are false or dangerous, justified or unjustified. Such judgments cannot initially be made by an individual but must stem from a collectivity; and because we all belong to communities, it makes sense for us to invoke such judgments. In sharp contrast, it makes no sense to indicate that a computer has made a mistake or is unjustified in its beliefs. The computer is simply executing what it has been programmed to execute, and standards of right and wrong do not enter into its performance. Only those entities that exist within, interact with, and are considered part of a community can be so judged.

My other reservation about the computer as model centers on the deep difference between biological and mechanical systems. I find it distorted to conceive of human beings apart from their membership in a species that has evolved over the millennia, and as other than organisms who themselves develop according to a complex interaction between genetic proclivities and environmental processes over a lifetime. To the extent that thought processes reflect these bio-developmental factors and are suffused with regressions, anticipations, frustrations, and ambivalent feelings, they will differ in fundamental ways from those exhibited by a nonorganic system. Note that it did not have to be this way—biological systems might have been just like inorganic (mechanical) systems. But it is clear that they are not. I therefore believe that adequate models of human thought and behavior will have to incorporate aspects of biological systems (for example, processes of organic differentiation or fusion) as well as aspects of mechanical systems (the operation of electronic circuits). The very comparisons between organic and mechanical structures and processes may be among the most instructive aspects of the science. All told, cognitive science will have to incorporate (and connect to) neurobiology as much as to artificial intelligence.
Central to my view of cognitive science is the claim that the field entails an empirical effort to answer long-standing epistemological questions. Classical philosophy has indeed supplied much of the intellectual agenda of the contemporary field: its interest in how we perceive the world, in how we classify objects, in the status of words, images, and other constructs, and in the assessment of human rationality or irrationality. And even those issues that could not have been formulated by the classical philosophers—such as the extent to which human thought is computational—have routinely been put forth in philosophical terms. Perhaps most notably, contemporary cognitive science has provided reasonable answers to selected philosophical questions, even as it has rejected certain issues and radically transformed others.

In a sense, philosophy can be seen as standing outside of mainstream empirical cognitive science. On one bank of the mainstream, philosophy supplies many of the issues to be investigated. On the other, it examines the answers that are forthcoming, helps to interpret and integrate them, and provides critiques of the overall enterprise. Thus, for example, philosophers first raised the issues of human rationality and have now participated vigorously and instructively in the interpretation of findings put forth by Freud, Piaget, Tversky, and Kahneman. Yet as members of a discipline that stands external to empirical science, philosophers concerned with cognitive science may seem in jeopardy. For, as philosophical questions are answered by empirical science, philosophers may ultimately recede from the scene—as has, in fact, happened in vast areas of physics and biology. Still, in my view, philosophers interested in cognition stand in no peril of intellectual unemployment for the foreseeable future.

But what of the relations obtaining among the other disciplines that make up the cognitive sciences? Are they likely to blend together into one seamless Cognitive Science, or can we expect them to maintain their autonomy in the years ahead? And what would be the most favorable state of affairs?

One can contrast two visions of cognitive science. The less ambitious one calls for cooperation among the six member disciplines, each still retaining its primary questions, methods, and goals as chronicled in part II. On such an account, philosophy supplies the principal issues and helps to judge the extent to which they have been successfully handled. Neuroscience and anthropology remain as border disciplines, psychology and artificial intelligence are the core disciplines, and linguistics offers an account of that ability which is most central in the human cognitive armamentarium. When collaborating, these researchers are "practi-..."
ing cognitive scientists”; otherwise, they are simply doing their own thing.

This “weak” version of cognitive science is quite possibly the norm today but scarcely warrants the label of an important new science. In a stronger, more gritty version of cognitive science, there will be gradual attenuation of disciplinary boundaries and loyalties. These will be replaced by a concerted effort by scientists committed to a representational account to model and explain the most crucial human cognitive functions.

This reconfiguration of the territory of cognitive science rests on the following analysis. Today, what is most central to cognitive studies is an individual’s disciplinary background: whether one works as a philosopher or an anthropologist is more salient than whether one works on issues of language or of social interaction. This organization around the traditional disciplines would be appropriate if the actual domains of cognition did not make a central difference; so long as the same processes are believed to occur irrespective of the content of a domain (musical versus spatial cognition, for example), the conventional disciplinary division of labor makes sense.

I hold a very different, and still controversial, vision. From my perspective, as elaborated in this book, the crucial divisions within cognitive science are not the traditional disciplinary perspectives but rather the specific cognitive contents. Therefore, scientists should be characterized by the central cognitive domain on which they work: broad domains like language, music, social knowledge, logical thought; and more focused subdomains, like syntactic processing, the early phases of visual processing, or the perception of rhythm. Scientific training and research enterprises should come increasingly to be organized around these problems. When working on these problems, scientists should fuse their necessarily different perspectives in order to arrive at a full account of the particular cognitive domain at issue. And so the ultimate cognitive-scientific picture of syntactic processing, or of language as a whole, should be a coordinated representational account which covers the full gamut of the traditional disciplines without any need even to mention them.

Yet the question of disciplines or, more broadly, of levels of explanation cannot be bypassed entirely—and here we confront the major challenge to contemporary cognitive studies. Having established the legitimacy of the representational level, cognitive workers must trace out the ways in which this level maps onto the other legitimate (and legitimized) ways in which human activities can be construed. For a time, believers in the representational level had to proceed along their own, as yet unexplored path—and adherence to this single-minded program was the genius of the pioneering generation of the 1950s. But ultimately, such splendid isolation must be shattered. We must come to understand how culture is mapped
onto brains—and the royal road toward such understanding will be the representational level.

The reason for such linkage across levels is simple but crucial. Unless the significance of work in each science can be connected to that undertaken in neighboring areas, the significance (and the limitations) of that work cannot be appreciated. No one fears the demise of physics, chemistry, and biology; and yet each of these disciplines has vital, articulated, necessary links to the next level, through "borderland" disciplines like physical chemistry or biochemistry.

But, paradoxically, much of the best work in cognitive science has been carried out as if only the level of mental representation existed. In the case of language (more specifically, grammar), for example, the brilliant work of Chomsky and his followers makes no reference to, and could be maintained irrespective of, the actual conditions in the brain and in the surrounding culture. If cognitive science is to mature, however, the ultimate representational account of language must relate, at one extreme, to knowledge about the neural architecture of certain regions of the left hemisphere of the brain; and, on the other, to knowledge about the structure and function of language in different cultural groups. Only such a linking of levels can indicate whether proposed representational accounts of language are in fact appropriate, in light of neural and cultural considerations. The goal of this penetration of levels is not, to repeat, so that one discipline or level can swallow the other; but rather, so that our understanding of a domain like language can touch on all the relevant scientific perspectives, from neuron to nation.

Just as many committed cognitive scientists have restricted their work to the representational level and have spurned the borderland territory, so too they have called for a narrow delineation of what counts as cognitive. Thus, Jerry Fodor (1981) has expressed skepticism about the capacity of cognitive science to explain any of the higher or more complex forms of thought, which are "permeable" to a person's beliefs; and Zenon Pylyshyn (1984) has proposed a definition of cognition which excludes areas like learning, development, and "moods."

Although perhaps a prudent research strategy in the short term, I find this a misguided overall program for the cognitive sciences. Just because our current measures or concepts are primitive, we ought not to violate a common-sense notion of what mind is about; even less should we want to bypass the most impressive achievements of the human mind.

Indeed, in my view, the ultimate goal of cognitive science should be—precisely—to provide a cogent scientific account of how human beings achieve their most remarkable symbolic products: how we come to compose symphonies, write poems, invent machines (including computers), or construct theories (including cognitive-scientific ones). Such accounts will
have to incorporate the means by which humans embark on complex projects to achieve ambitious goals; how they represent their plans; how they initiate work on a project, organize their daily routines (and nonroutines!), evaluate tentative drafts in light of feedback from other people and in view of their own motives and standards, determine when such a program or product has been completed, and then initiate a new line of work.

Such an exploratory enterprise will probably entail mentalistic entities or models of a highly molar form (as well as many finer-grained entities). It will probably cut across narrow domains (like language-related processes) and also have to involve separate constructs to account for processes involved in creativity, synthesis, and/or consciousness. Ultimately, as part of the cognitive challenge, it will also be necessary to relate a representational account of these human intellectual achievements to what is known about their neural substrate and to what can be established about the role of the surrounding culture in sponsoring and then absorbing (or rejecting or refashioning) them.

Even to begin to outline the phases involved in modeling any complex human creative activity is to confront the immensity of the task and the primitiveness of our current tools. And yet it is crucial for cognitive scientists to keep this goal in mind—even to tack it over their desks or alongside the screens of their personal computers. The study of thought must not exclude its most remarkable exemplars even if their elucidation still seems remote.

Given the most optimistic scenario for the future of the cognitive sciences, we still cannot reasonably expect an explanation of mind which lays to rest all extant scientific and epistemological problems. Still, I believe that the authors of *Meno, The Discourse on Method, The Critique of Pure Reason,* and *The Origin of Species* would feel that distinct progress has been made on the age-old issues that exercised them. Thanks to the development of new logical tools, the diverse deployments of the computer, the application of the scientific method to human psychological processes and cultural practices, our deeper and more rigorous understanding of the nature of language, and the many discoveries about the organization and operation of the nervous system, we have attained a more sophisticated grasp on the issues put forth originally by Plato, Descartes, Kant, and Darwin.

How much further cognitive science can proceed, and which of the competing visions it will choose to pursue, are issues that remain open. All who style ourselves as cognitive scientists are on the spot. If we heed the lessons entailed in our scientific history and lurking in our philosophical backgrounds, if we attend to but are not stymied by the reservations aired by shrewd skeptics, if we recognize the limitations of all inquiry but do not thereby encounter a failure of nerve, there are clear grounds for optimism.
Epilogue to the Paperback Edition

Cognitive Science
After 1984

Cognitive science is perhaps forty years old as a concept, and ten as a self-described scholarly discipline. Despite its brief life, the field has already gone through a number of phases, which I have described in the previous chapters. In the even briefer period since this book was completed, new trends have begun to emerge—upon which it is apposite to touch as The Mind's New Science appears in a paperback edition.

In reviewing the past few years, I am struck by a paradox. The kinds of problem on which leading-edge scientists are currently working, and the ways in which they are approaching them, bear a much closer relationship to the Weltanschauung of scientists on the eve of the founding of cognitive science than to work carried out by their successors a decade or two ago.

Let me try to unpack this paradox. During the 1940s, the "forefathers" of cognitive science had grown increasingly impatient with a behaviorist approach, which avoided discussion of the brain, rejected conceptions of mental representation, and averted consideration of higher-level perceptual or problem-solving processes. Scientists like W. Ross Ashby, Donald Hebb, Karl Lashley, Warren McCulloch, Heinz von Foerster, John von Neumann, and Norbert Wiener articulated a contrasting vision. As they saw it, our understandings of the brain and of the nature of computation could come together in the study of cognitive systems—and particularly

I thank Mark Bickhard, Hiram Brownell, Judy Greissman, Phoebe Hoss, Stephen Kosslyn, Stephen Pinker, Eric Wanner, Ellen Winner, and Edgar Zurif for their help with this essay.
of those exhibited by the human mind. It was this vision—admittedly in varying forms—that underlay Warren McCulloch and Walter Pitts's modeling of neural networks in terms of the logical calculus; Donald Hebb's neuronal models of early perception and learning; Frank Rosenblatt's somewhat later, ill-fated attempts to simulate recognition via perceptrons; and the broader synthesizing accounts put forth by Lashley, von Neumann, and Wiener.

While never completely forgotten, this vision was at least tentatively suspended by the first generation of workers in cognitive science. Whatever their differences, pioneering cognitivists like Jerome Bruner, Noam Chomsky, Claude Lévi-Strauss, John McCarthy, George Miller, Jean Piaget, and Herbert Simon showed little concern with the human brain or nervous system. Their focus fell almost entirely on the properties of thought, as suggested by systematic introspection, probed in the psychological laboratory, simulated by the computer, or embodied in the symbolic products of a culture. They made—and won—the case for a level of mental representation, the modeling of higher cognitive processes, and the surface compatibility between human and computer symbolic systems.

The scientific pendulum has swung in a different direction in the past few years. As the limits of the serial digital "von Neumann computer" became clearer, and as alternative views of computation gained in persuasiveness, there has been a shift to a different "modal view" of cognition—a view in which psychological, computational, and neurological considerations are far more intricately linked. This point of view was already described briefly in chapter 10 (pages 318 and 321) and in the conclusion (pages 388-92) of this book. Within the last few years, the principal claims and the broad ambitions of this "parallel distributed processing" approach have become even clearer.

To begin with, the PDP approach (as it is frequently abbreviated) styles itself explicitly in opposition to the classical Newell–Simon and the Fodor views of cognition. Instead of serial operations or computations upon symbols or strings of symbols, instead of "executives," "interpreters," and "central control units," the PDP approach typically posits thousands of connections among hundreds of units (in principle, the approach can be extended to millions or even billions of connections). The resulting networks feature the signaling of excitations or inhibitions from one unit to another. "Perception," "action," or "thought" occur as a consequence of the altering of the strengths (or weights) of connections among these units. A task is completed or an input processed when the system ultimately "settles" or "relaxes" (at least tentatively) on a satisfactory set of values or "stable states"—in short, upon a "solution."
This family of models has been shown to work quite neatly with a range of problems drawn deliberately from diverse areas of perception and thinking. PDP models are capable of simulating finger movements in skilled typing, the reaching for an object without falling over, and the perceptual "completion" of familiar patterns or illusions. These models can also handle problems that arise at a more conceptual level. They can learn, for example, to distinguish among fifty-five different animal species on the basis of the features that characterize these animals; or to ascertain the organization to which a target individual belongs, on the basis of demographic information about membership in a number of organizations.

I find particularly intriguing James McClelland's demonstration that a PDP model is able to acquire the procedures whereby the past tense is formed in English; and that the model initially makes the same errors of overgeneralization of the "-ed" ending as do young children. Such findings hold promise for an ultimate convergence between computational and developmental accounts—a convergence that had seemed remote under the classical cognitive-scientific view.

Because, in PDP models, all knowledge inheres in the connections themselves, proponents have offered a reconceptualization of major psychological faculties. Thus, instead of being viewed as a set of facts or events stored in the brain, memory is viewed as the set of relationships that obtain among various aspects of facts or events as they are encoded in groupings or patterns of units. What is stored are the connections and strengths among units which allow the patterns to be subsequently re-created. Analogously, learning is a matter of finding the right connection strengths so that proper patterns of activation are produced under the appropriate circumstances.

As should be evident, the PDP models operate in a manner more reminiscent of the human brain. It has always been conceded, even by the most ardent defenders of the classical cognitive view, that the brain has to operate at least in part via parallel processing: but this fact was not considered damaging to the prevalent serial view. It must therefore be stressed that the PDP views do not claim to be actual models of brain processing: as David Rumelhart and James McClelland indicate, they are not devising neural models but are rather engaged in "neurally inspired modelling of cognitive processes" (1986, vol. I, p. 130). It nonetheless must be gratifying—and perhaps also significant—that the kinds of process that characterize PDP models appear to resemble those thought to occur in the mammalian nervous system.

I also note two other potential convergences among disparate levels of cognitive science. At the level of neuronal analysis, neurobiologists like
Eric Kandel have long maintained that learning consists in an altering of strength among pre-existing connections, rather than in a construction of new synaptic connections. Not only does this point of view mesh nicely with current artificial intelligence work of the PDP variety; but it is also echoed in Noam Chomsky's recent statements about how a particular language is acquired (see page 211 of this book). Chomsky now views the newborn infant as possessing an intricately wired system, some of whose connections have not yet been established. The switches or parameters may be set in a limited number of ways. As the young child is exposed to language, these parameters are set by experience; and once they have been set in a permissible way, the individual possesses a particular language—with its unique options and limitations.

And there are reverberations as well with basic scientific work by Nobel laureate Gerald Edelman on the nature of learning systems at the cellular and neurochemical level. In Edelman's view, the pattern of neural connections is fixed at birth, but certain combinations of connections are selected over others as the result of particular stimuli encountered and of resultant competition among different neuronal groups. The different groups of cells, or "maps," speak to one another to create particular categories of things and events, which themselves are altered over time by subsequent experiences. Edelman offers his model as a way of accounting for both the universal and the highly individualized aspects of human cognitive experiences.

Philosophical and psychological circles have also witnessed a reaction against the assumption that cognition necessarily consists in the manipulation of symbols and in the following of explicit rules. From the philosophical perspective, Stephen Kosslyn and Gary Hatfield maintain that it is possible to put forth an account of perception (and possibly other cognitive capacities as well) in terms of representations, without any need to evoke symbols, operations upon them, or rules stored and followed by a system. The requisite processing can be accomplished by the design of the system as such. Under this analysis, representations are simply the states of an organism or a device that qualify as representations by virtue of the role they play in the performance of a task or function proper to that organism or device.

From a complex developmental and epistemological perspective, Mark Bickhard and his collaborators also argue against an account of cognition which rests on symbolic encoding. While the concept of representation must ultimately find its place in any account of cognition, accounts based upon encoding lead to an infinite regress, for one must always have a prior meaning to be encoded. As Bickhard explains, "Either x must
represent (encode) \( y \), or \( x \) must encode whatever it represents—which does not succeed in providing any representational content to \( x \), and, therefore, does not succeed in making \( x \) an encoding at all” (personal communication, 1986).

Bickhard puts forth an alternative point of view. He believes that, à la Piaget, we must think of organisms as perpetually interacting with an environment in pursuit of various goals and competences. Representations emerge in development as the consequence both of differentiations of the environment and, more generally, of the operation of certain kinds of goal-directed interactive control structures. Within this perspective, encodings may be derived as specializations of the emergent representational function.

These recently articulated philosophical and developmental perspectives share at least some affinity with the Gibsonian approach to perception. However, their recognition of the necessity for some kind of representational account places them squarely within, rather than outside of, the cognitive-scientific tradition.

Of course, in the face of these shifting currents, those associated with the “classical approaches” in cognitive science have not completely abandoned their previous positions. The issues of inferencing and intentionality, raised by Fodor and Pylyshyn in their critique of Gibson’s position (pages 311–13), can be resurrected with respect to PDP approaches. Stephen Pinker and Alan Prince exemplify this skepticism in their critique of PDP models of linguistic processes. These critics point out that, by dealing with sheer “sounds” rather than “phonological units” (the sounds of language) or “morphological units” (words), the PDP modelers are compelled to miss important generalizations about language. A coherent and comprehensive account of linguistic processes and categories must involve a number of levels of representation and incorporate adequate ways of representing linguistic input and output; left to their intrinsic self-organizing principles, PDP accounts cannot adequately deal with linguistic phenomena.

Even those sympathetic to PDP approaches concede that they operate more effectively with perception and other “lower level” (subsymbolic) processes than with large-scale problem-solving, problem-finding, invention, and other “symbolically laden” enterprises. As Rumelhart and his colleagues succinctly put it, what is difficult to describe in the PDP framework are “the process of thinking, the contents of consciousness, the role of serial processes, the nature of mental models, the reasons for mental simulations, and the important synergistic role of language in thinking and in shaping our thought” (1986, vol. II, p. 38).
It may be, as I suggested in the hardback edition of this book, that models of human cognition will ultimately have to incorporate aspects of both the PDP and the classical "serial-symbolic" models. Lower-level perceptual processes can be accounted for adequately in terms of PDP models, but higher-level conceptual or linguistic processes (and perhaps even some aspects of perception) will require a more complex, multilevel representational account. I predict that, over the next decade or two, the pendulum will swing back and forth between these two contrasting approaches. If we are fortunate, each swing will involve greater and more sophisticated integration among the approaches.

In describing the forces that contribute to the creation of new knowledge in a field like cognitive science, I have found it useful to adapt a scheme developed by Mihaly Csikszentmihalyi, in collaboration with David Feldman and me. We distinguish among three components, each of which figures inextricably in significant creative activity. There are the individuals, possessing various skills and talents, who choose to work on a set of issues. There is the surrounding field, the ensemble of institutions, organizations, roles, publication outlets, and so on, which selects the contributions of certain individuals as particularly notable. Finally, there is the domain, the organized structure of knowledge in a discipline or activity. This structure alters over time, and the revised domain becomes the body of knowledge encountered by the next generation of workers in a discipline.

The hardback edition of The Mind's New Science began with a description of some of the principal individuals involved in the launching and flowering of cognitive science. Viewed a few years after publication, the list would be essentially unchanged; perhaps the major exception would be greater attention to those individuals most closely associated with the parallel distributed processing approach, whose findings are presented in the Rumelhart and McClelland volumes. I would also credit important work in the 1940s by the Austro-American cyberneticist Heinz von Foerster and the British polymath Kenneth Craik.

As for the emergence of the field of cognitive science, I am pleased to report that its reputation and its scope have risen more rapidly than I would have predicted. In June 1985, representatives from approximately fifty colleges and universities assembled at Vassar College to discuss undergraduate education in cognitive science. In the last year or so, several major universities—among them Brandeis University, Brown University, MIT, and UCLA—have reorganized their graduate education so as to take into account the growing importance of interdisciplinary work in the cognitive sciences. I would estimate that over one hundred colleges and uni-
Universities in the United States and Canada now have recognized programs in the cognitive sciences, with that number increasing almost monthly. Perhaps equally telling, many individuals now label themselves as "cognitive scientists" and that appellation is no longer unfamiliar in technical writing and in more "up-scale" general publications. Of course progress in the field cannot be automatically equated with progress in the domain, but it may well help to set the stage for scientific developments.

The domain of cognitive science consists of the accruing of knowledge about human cognition, through disciplinary and (increasingly) interdisciplinary scholarship. In the short run, such progress is difficult to assess with confidence. My own judgment is that the research program described in The Mind's New Science is being actively pursued on many fronts: a substantially changed second edition could be written, describing significant advances in our understanding of visual perception, natural language processing, imaging, categorization, and human rationality. In lieu of such a revision, it is appropriate to conclude this epilogue with an annotated listing of those texts that have proved most useful in understanding cognitive science after 1984.


The first two "gospels" on parallel distributed processing models:


Critical updates on artificial intelligence:


Epilogue to the Paperback Edition

A compendium of recent studies in cognitive anthropology: The introduction describes "cultural models" and "cultural programs" of a far more general sort than "scripts" and "frames"; Holland, Dorothy; and Quinn, Naomi. In press. Introduction to Cultural Models in Language and Thought. Cambridge: Cambridge University Press.


Attempts to present a representational view that does not incorporate encoding or symbols:


Howard Gardner
Cambridge, Massachusetts
January 1987