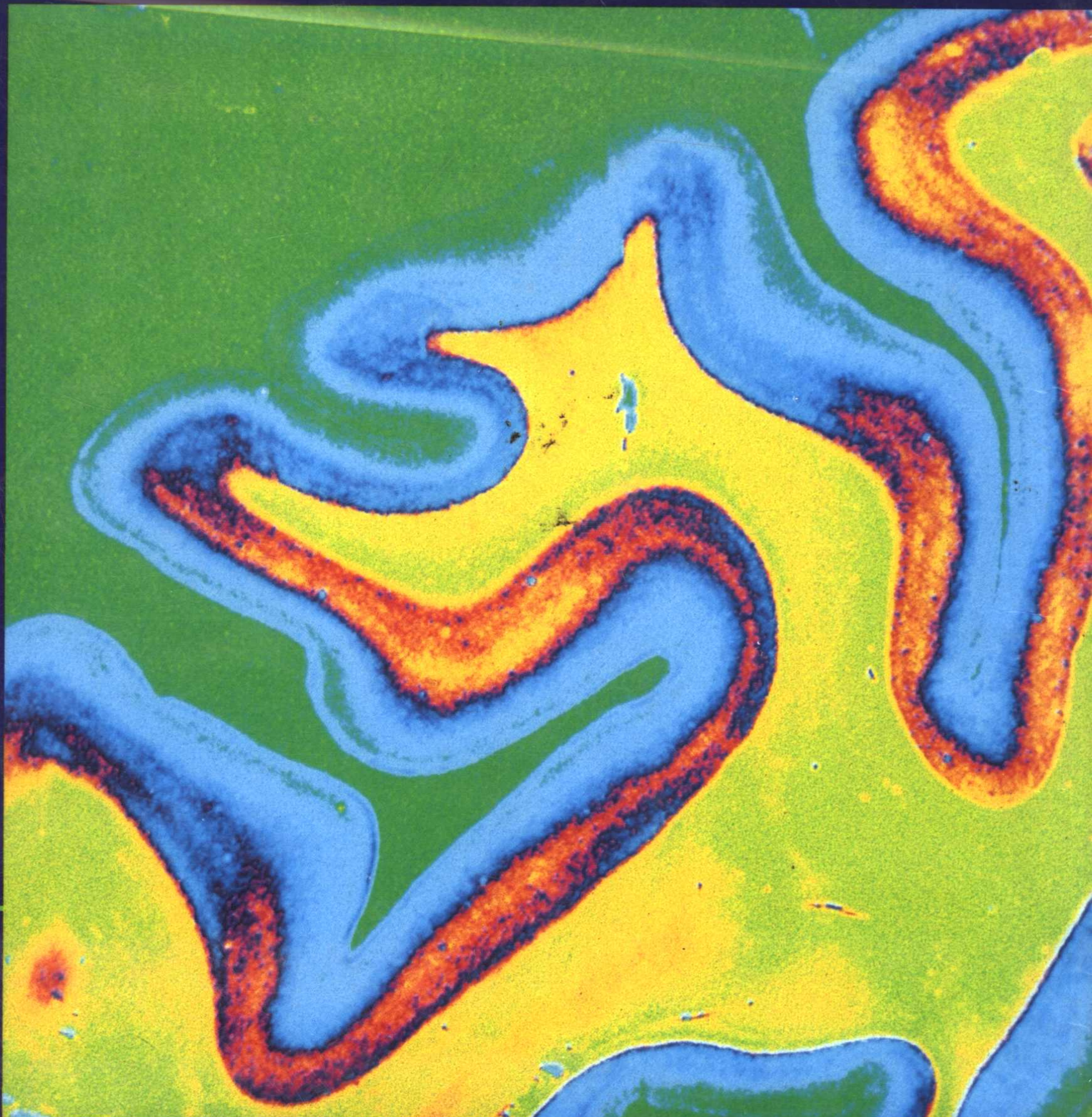
FRONTIERS IN COGNITIVE NEUROSCIENCE



edited by
Stephen M. Kosslyn
and Richard A. Andersen

Frontiers in Cognitive Neuroscience

edited by Stephen M. Kosslyn and Richard A. Andersen

A Bradford Book

The MIT Press
Cambridge, Massachusetts
London, England

This book was printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Frontiers in cognitive neuroscience/edited by
Stephen M. Kosslyn and Richard A. Andersen.
p. cm.
"A Bradford book."
Includes bibliographical references and index.
ISBN 0-262-11163-2
1. Cognitive neuroscience. I. Kosslyn, Stephen Michael,
1948– . II. Andersen, Richard A.
QP360.5.C64 1992
153—dc20 91-36146
CIP

General Introduction
Stephen M. Kosslyn and Richard A. Andersen

It has sometimes been said that "the mind is what the brain does." Cognitive neuroscience is built on this assumption, and its goal is to understand how brain function gives rise to mental activities such as perception, memory, and language. To understand the excitement surrounding the emergence of cognitive neuroscience, we must see how it arose from the confluence of several fields.

Cognitive neuroscience rests in part on the idea that different parts of the brain do different things—an idea with a checkered history. To understand the cognitive neuroscience approach, and why the time is finally ripe for this enterprise, we will consider first its historical roots and then turn to the factors that led to its recent emergence.

Historical Roots

Although philosophers have always been interested in the mind, it was not until the nineteenth century that scientists began to consider in earnest how mental activity arises from the brain. The major issue was whether the brain functions as a single, integrated whole (as the "globalists" believed) or as a collection of distinct organs, each responsible for a separate ability (as the "localizationists" believed). This central debate involved several lines of research during this period. Three examples are the phrenologists' view of localization of mental abilities, Pierre Flourens's and Gustav Theodor Fritsch and Edvard Hitzig's rather different conclusions based on experiments in animals, and Paul Broca's clinical findings of very specific deficits following brain lesions in humans.

The phrenologists, led by Franz Joseph Gall and J. G. Spurzheim, developed a "faculty psychology," which claimed that different regions of the brain implement distinct mental faculties, a not-so-unreasonable hypothesis. They also claimed, unfortunately, that more accomplished abilities are carried out by larger regions of the brain and that larger regions of the brain in turn create larger bumps on the skull. Neither assumption proved correct. Moreover, phrenologists characterized these faculties using rather arbitrary categories of human behavior and experience, which led them to assume that faculties such as sublimity, secretiveness, and parental love are carried out by distinct regions of the brain. This also proved to be off the mark.

Although phrenology was deeply flawed, it also produced some important concepts. For example, the phrenologists were the first to emphasize the importance of the surface of the brain, the cortex; previous theorists focused on the interior of the brain (typically concentrating on the role of the ventricles). In addition, phrenologists stressed that the brain is not a single, undifferentiated system, which also turns out to be correct.

Perhaps the most serious shortcoming of the phrenologists was that they were not really scientists; they did not put their ideas to rigorous empirical test. Instead, they felt bumpy skulls and made claims about the owners' personalities and mental abilities without trying carefully to validate their claims. Although phrenology was a pseudoscience, it did develop a clear set of ideas that could be treated as hypotheses by working scientists, who took up the challenge. These early researchers focused on the claim that mental abilities are localized to specific brain sites, rather than on the ancillary conceptions of what the abilities are and how to assess them. In these studies physiologists examined the effects of lesions or stimulation of different brain areas on the behavior of animals.

Pierre Flourens (1794–1867) is usually cited as one of the first physiologists to spearhead an attack against the localizationist view by using animal experiments. Flourens performed ablation studies on birds, removing only selected parts of their brains and then observing their behavior after they recovered. He claimed that the birds recovered, regardless of where the brain was cut, and regained the same set of abilities. These findings led Flourens to argue that the brain works as an integrated whole and that specific abilities are not localized to particular sites.

In retrospect, Flourens's research was flawed in several critical respects. First, his behavioral tests were crude. These methods were so insensitive that Flourens probably would not have noted selective deficits even if they were present. Indeed, modern researchers now report selective effects of damaging specific locations of birds' brains (e.g., Nottebohm 1981). Second, one must wonder about the precision of his surgery; it seems likely that his lesions often produced widespread damage, and so they may not have selectively affected specific neural structures.

Moreover, other scientists did produce evidence for the localization of function. Perhaps the best example is the research of Fritsch and Hitzig (1870), who delivered electrical stimuli to the cerebral cortex of dogs. Fritsch and Hitzig discovered that shocking specific locations of the cortex caused certain muscles to twitch. These results supported the idea that at least some functions are localized to distinct sites in the brain.

The third facet of the debate took place at about the same time that the animal experiments were underway. In contrast, however, this aspect of the debate focused on clinical phenomena. This third set of researchers were primarily physicians, who observed selective behavioral deficits when their patients suffered brain damage. Although others predated him, Broca's 1863 report probably had the most impact on the debate.

Broca described a patient who had suffered a stroke. A stroke occurs when blood is cut off from part of the brain, which causes those cells to die. This patient had a severe language deficit following his stroke; indeed, after the stroke all he could say was "tan." Broca examined the patient's brain immediately after he died and found that the stroke had damaged the posterior part of the left frontal lobe. Broca reported a second case with similar damage later that year. After seeing a number of such cases, Broca inferred that language depended on specific folds in the cortex, specifically the left posterior portion of the third frontal convolution.

The localizationist view received even stronger support when, in 1874, Carl Wernicke reported that a different language deficit occurred when a different part of the brain was damaged. His patient had damage to the left superior temporal gyrus and had trouble not in producing language—as had Broca's patients—but rather in comprehending language.

It was not long before others reported a wide variety of behavioral deficits that were correlated with damage to different brain sites. These reports led to the "era of the diagram makers." These investigators constructed diagrams that specified the locations of various brain sites for specific behaviors and described how these sites are anatomically connected. These early diagrams were often based on superficial analyses of the behavioral deficits and on the suspected localization of the damage. Thus, it was not surprising that many viewed them with skepticism.

At about the same time that these clinical findings were being reported on the Continent, the British neurologist John Hughlings Jackson offered a more sophisticated view (which in some ways was similar to Wernicke's views), suggesting that multiple brain areas contribute to complex brain processes. Although Jackson discovered that damage to the right cerebral hemisphere impairs visual perception more than damage to the left hemisphere, he rejected the idea that such a complex cognitive function can be carried out solely in one particular location of the brain. Rather, Jackson found that particular functions were never completely lost after a stroke; in fact, in some contexts patients could often exhibit abilities that were impaired or absent in other contexts. For example, it is not uncommon to find patients who cannot touch their lips with their tongue when asked to do so, but have no trouble performing this act to remove a crumb. Indeed, some patients cannot say specific words, such as "no," unless they are emotionally agitated. These sorts of observations led Jackson to claim that multiple regions of the brain are involved in carrying out complex functions such as perception, action, and language (see Taylor 1932).

The Twentieth Century

Remarkably, this debate continued into the twentieth century, in spite of the rather strong data for localization of function that existed in the experimental and clinical literature. The turn of the century marked the Golden Age of neuroanatomy. Santiago Ramón y Cajal, a Spanish neuroanatomist, using a stain developed by his contemporary (and competitor) Camillo Golgi, documented the immense complexity and anatomical specificity of the brain during the last part of the previous century and the early part of this century. Of particular interest is Cajal's theory that the brain is made up of individual elements, neurons. This idea was in contrast to the views of Golgi, who believed that the neural plexus was a continuous structure, a syncytium. Another insight of Cajal's was that nerve cells are unipolar, collecting information in the dendrites and sending it out along the axon. These basic ideas are some of the building blocks that began a century of progress in our understanding of brain function.

At the turn of the century, the famous German neuroanatomist Korbinian Brodmann used a stain originally developed by Franz Nissl (which stains the endoplasmic reticulum in the cytoplasm of cells) to study differential staining in the cortex. He used cytoarchitecture, the variation in appearance of various cortical areas due to different packing densities and cell morphologies, to distinguish roughly 50 areas. A large number of cytoarchitectural studies emerged in the period following Brodmann, including those by the Vogts, von Economo, von Bonin and Bailey, and Walker, which further divided the cortex into many cortical areas. Some of the most obvious divisions have remained, such as the border dividing Brodmann's area 17 (striate cortex) from Brodmann's area 18. Many other divisions have not remained; for instance it is now known that area 18 contains several different visual areas. Although perhaps a majority of the subdivisions recognized by the cytoarchitectonic studies have not stood the test of time, the paper by Merzenich and colleagues in this collection is a small victory for Brodmann. Areas 3a, 3b, 1, and 2 of Brodmann had been lumped in subsequent years by physiologists into one somatosensory area, SI. Merzenich and colleagues showed that there are, in fact, at least three representations of the body surface within SI, which correspond to Brodmann areas. Another more recently exploited technique involves staining the myelin sheaths that surround the axons of nerve cells; the use of myeloarchitecture has proved valuable in distinguishing area MT, the visual motion area whose function is described in the papers by Allman, Miezin, and McGuinness and by Movshon, Adelson, Gizzi, and Newsome in this collection.

The discovery in the earlier part of this century that the brain has a complex physical structure seemed to many to be strong support for the localizationist view. Some researchers felt at that time (circa 1930), however, that differences in appearance did not necessarily imply differences in function, and others questioned whether these structural differences existed at all. In particular, an influential psychologist, Karl Lashley, and his colleagues not only challenged many of the claims about the physical variations within the cortex (Lashley and Clark 1946), but also performed Flourens-like experiments on rats and reported similar findings. These findings were consistent with two principles Lashley (1929) used to explain the effects of brain damage on behavior. The principle of *mass action* stated that the brain operates as a single, integrated system, and the principle of *equipotentiality* stated that all portions of the brain have equal abilities. Taken together, these principles imply that the sheer amount of the brain that is damaged determines the behavioral deficit, not which particular parts are damaged.

Although Lashley's experiments were methodologically superior to those of Flourens, they were still crude by contemporary standards—and in fact failed to reveal selective deficits that have since been reported. Fortunately this brief regression was followed by an explosion of important discoveries regarding functional specialization in the cortex, and Lashley's influence was mitigated.

An important step forward in neurophysiological research occurred in the late 1930s when Woolsey, Bard, Rose, and their colleagues began novel experiments in which they mapped the evoked activity for sensory stimuli by placing wick electrodes on the surface of the cortex of animals. They mapped two auditory areas, AI and AII, and two somatosensory areas, SI and SII. As mentioned above, it was later found that SI contains several distinct areas, and their area AII contains three areas. Based on their recordings from visual cortex, they also proposed two areas; although there is a certain symmetry to their initial studies, it is now appreciated that there are at least 30 visual areas in the monkey cortex.

The surface electrodes used by these investigators recorded the composite activity of many thousands of nerve cells. With the advent of single-cell recording, using very fine wires insulated except at their tips, John Allman and Jon Kaas (reviewed in Woolsey 1981) began to map the visual cortex in new-world monkeys. They found one representation of visual space in the primary (striate) visual cortex, a fact that had been known for many years based on scotomas (blind spots) experienced by humans with brain lesions and the surface electrode technique with animals. When they continued to move into visual cortex outside of the primary visual cortex, however, they found several more representations of visual space, corresponding to several other visual fields. At this time Semir Zeki began tracing connections between the primary visual cortex and the extrastriate cortex and found that there were projections to several areas. This work is reviewed in Zeki's paper in this collection. Zeki also made the seminal finding that cells in area MT (V5) preferred moving visual stimuli, and cells in area V4 were selective for color. These important experiments laid the foundation for an exciting two decades of discovery, providing evidence for separate processing streams within the visual cortex. Much of this work is summarized in the papers by DeYoe and Van Essen, Livingstone and Hubel, and Schiller and Logothetis in this volume. At the same time a similar set of investigations was discovering multiple areas in the auditory cortex (reviewed by Brugge and Reale) and in the somatosensory cortex (Merzenich, Kaas, Sur, and Lin).

These empirical studies have led to a simple, elegant solution to the globalist/localizationist controversy. The brain is precisely organized, more so than any of the early researchers ever imagined. For instance, there is a small subdivision of the medial superior temporal area, about 3 mm in diameter, which in every monkey contains receptive fields that are selective for patterns of motion, such as rotation or expansion, occurring anywhere in their large visual fields. Similarly, circuits that perform specific transformations of the sensory input are found in striate cortex (Bolz, Gilbert, and Wiesel 1989). The mistake of the early localizationists is that they tried to map behaviors and perceptions into single locations in the cortex. Any particular behavior or percept is produced by many areas, located in various parts of the brain. Thus, the key to resolving the debate is to realize that complex functions such as perception, memory, reasoning, and movement are accomplished by a host of underlying processes, each of which confers only a single facet of the ability. These processes are relatively simple and mechanical; they do not "think" but rather reflexively perform a specific operation when provided with appropriate input. It is these simple, underlying processes that are carried out in a single region of the brain. Indeed, the abilities themselves typically can be accomplished in numerous different ways, which involve different combinations of processes. An example of this new synthesis can be gleaned from the readings in the section on attention, which include discussions of how the behavior of attending invokes the activity of many cortical areas (for example, Crick; Moran and Desimone; Richmond and Sato; Wurtz, Goldberg, and Robinson). Mesulam points out in his article, however, that lesions in humans to different attention-related brain structures produce different types of attentional deficits, leading him to suggest that the attention system is represented by a distributed network with its different nodes accomplishing different aspects of the process.

Any given complex ability, then, is not accomplished by a single part of the brain. So in this sense, the globalists were right. The kinds of functions posited by the phrenologists are not localized to a single brain region. However, simple processes that are recruited to exercise such abilities are localized. So in this sense, the localizationists were right (for further discussion, see Kosslyn and Koenig 1992; Luria 1980; Squire 1987).

Thus the stage was set for cognitive neuroscience: Two central questions emerged from the efforts of previous researchers. First, what are the simple, underlying processes that are carried out by the brain? And second, how do these processes work together to produce "mental" abilities? Both questions led to a new way of conceptualizing brain function.

The Emergence of Cognitive Neuroscience

Cognitive neuroscience arose when researchers conceived of brain function from a new perspective. This perspective grew out of a confluence of discoveries and ideas in three older disciplines, namely neuroscience (specifically neuroanatomy and neurophysiology), experimental psychology, and computer science. Although there are many modern precursors to the field (e.g., see Grossberg 1987; McCulloch and Pitts 1943), cognitive neuroscience probably did not truly come into its own until the late 1970s.

One foundation of cognitive neuroscience emerged in large part from the work of David Hubel and Torsten Wiesel at the Harvard Medical School and Vernon Mountcastle at Johns Hopkins University. These researchers led the way in helping us to understand how specific neural events could give rise to perception. They not only began to characterize elementary processes that are carried out by individual neurons but also discovered key facts about how the brain is organized. They found that some neurons are tuned for specific types of stimuli, that neurons are organized into "columns" with an orderly internal structure, and that these columns are interconnected in orderly ways. By combining neurophysiology (the study of dynamic properties of brain cells) and neuroanatomy (the study of the structure of the brain), these researchers and their contemporaries made great strides in showing how specific neural processes might be neurally integrated to produce sensory percepts.

Another important advance in neuroscience came with the melding of psychophysics with neurophysiology. Psychophysics is not the product of a deranged physicist but rather the study of the limits and parameters of human perception and ability. It represents an early branch of experimental psychology whose roots are the work of some of the great psychologists of the nineteenth century, and it has undergone a recent rebirth since the advent of computers, which enable highly precise stimulus control and response measurement. Psychophysical results tell us what the neural machine can do—its specifications, if you will—and give neurophysiologists clues about what to look for when deciphering brain mechanisms. An example is the study of depth perception. Bela Julesz, then at Bell Labs, and Gerald Westheimer, at the University of California at Berkeley, studied the abilities of humans to see depth from stereopsis. Their findings, and a unique stimulus developed by Julesz (the random-dot stereogram) enabled physiologists such as Gian Poggio, at the Johns Hopkins Medical School, to search for and find the neural substrate of this perceptual process in the monkey visual cortex.

An important technical advance in neuroscience was the advent of recording the activity of single nerve cells in behaving animals. This technique was first exploited in a systematic fashion by Edward Evarts at NIH. Evarts studied the motor cortex, which can only be appreciated when an animal can make voluntary movements. Soon after Evart's initial studies, David Robinson and his colleagues at the Johns Hopkins Medical School began using this technique to study the control of eye movements in monkeys. The eye movement system is a much simpler system than the limb movement system and as a result has proved to be an important model system for the study of motor control. Emilio Bizzi and his colleagues at the Massachusetts Institute of Technology further advanced this field by studying the coordination of combined eye and head movements.

Another foundation of cognitive neuroscience grew out of experimental psychology. Although there were a vast number of important contributors, three played a special role. Saul Sternberg, at the University of Pennsylvania, developed a ground-breaking technique for characterizing individual mental operations. Sternberg (1969) showed that some tasks are accomplished by a series of discrete processing stages and developed a method for characterizing the individual stages. Each stage was understood in terms of information that is stored and operated on in specific ways. Although his method has since proven less straightforward than originally conceived (because many tasks are not performed by a series of discrete serial operations), Sternberg's work helped to focus experimental psychologists on the problem of measuring individual processes that underlie complex abilities. One consequence of this analysis was the development of many sensitive behavioral tasks, which subsequently have allowed researchers in cognitive neuroscience to assess the effects of brain damage and to correlate local brain activity with specific types of information processing.

Another key contributor was Michael Posner, at the University of Oregon. Posner developed tasks that assess relatively simple aspects of information processing (e.g., see the paper reprinted in this volume), and developed conceptual foundations for the study of attention and related abilities. Like Sternberg, Posner developed ways of testing individual components of information processing, but his ideas were not tied to specific models of sequential processing. His elegantly simple tasks and sophisticated ideas have subsequently played a major role in the study of effects of brain damage on behavior and in the use of brain-scanning techniques to study the localization of cognitive function.

Much research in experimental psychology stressed the discrete nature of cognitive operations. This point of view was nicely counterbalanced by the work of Roger N. Shepard, at Stanford University. Shepard used behavioral data to demonstrate that the brain performs "analog" operations, such as mentally rotating or folding objects in visual mental images. For example, he and his colleagues found that subjects required progressively more time to imagine an object rotating through greater arcs or being folded more times (see Shepard and Cooper 1982). This work also demonstrated that there must be internal representations of images in the brain and challenged many researchers to gain an understanding of the properties of different kinds of internal representations. One of the papers in this collection has taken up this challenge. Georgopoulos and his colleagues report what appears to be a neural correlate of a mental rotation of a planned movement.

The idea that behavior can be understood in terms of operations on internal representations led many experimental psychologists into cognitive science. Cognitive science likens the mind to a computer program, which has led researchers to try to characterize the representations and operations carried out by the mind. This new brand of experimental psychology has played a special role in the development of cognitive neuroscience, emphasizing the structural organization of information processing.

Another important development in experimental psychology was the widespread realization that behavioral data alone are not sufficient to characterize mental processes. John Anderson (1978) proved that any set of behavioral data could always be explained by more than a single theory and suggested that neurophysiological constraints would help to ameliorate this problem. This approach led some experimental psychologists to look more seriously at the neural substrate of behavior.

The third foundation of cognitive neuroscience was provided by developments in computer science—in particular in the subdiscipline of artificial intelligence (AI). The reconciliation of the globalist/localizationist debate led researchers to expect relatively simple processes to be localized in specific regions of the brain; however, researchers in neuroscience characterized brain function largely on the basis of common sense. For example, if a single cell in the visual cortex responded vigorously to an oriented line segment, it might be construed as an "oriented-line detector." Unfortunately, many neurons respond in complex ways to stimuli, and common sense falters when trying to characterize their function.

Computer science provided a new, more powerful way to think about brain function. From the start, researchers in AI were confronted with the problem of characterizing elementary processes. Their goal is to build machines that behave "intelligently," and they typically build such machines by programming hierarchies of processes. That is, they compose complex functions out of sets of simpler ones. Indeed, Herbert Simon, one of the founders of AI, argued that all complex devices should be built in this way (Simon 1981).

The simple processes in AI programs execute individual *computations*. The concept of a "computation" is closely related to the concept of a mapping; a computation systematically maps an input to an output. Moreover, both the input and the output convey information, and the relation between the two can be characterized in terms of a (mathematical) function. Computations transform inputs to produce appropriate outputs.

Because a computation describes a type of mapping, this concept could be applied equally easily to a computer or to a brain—as the early researchers in cybernetics, who focused on analog computing systems, concluded. Indeed, John von Neumann (1958), Norbert Wiener (1948), Warren McCulloch and Walter Pitts (1943), and others regarded neural processes as performing computations.

Many researchers adopted this conceptualization of brain function after learning about the work of one man, the late David Marr of the Massachusetts Institute of Technology. At the time, Marr's work was uniquely interdisciplinary and was particularly important because it provided the first rigorous examples of cognitive neuroscience theories (Kosslyn and Maljkovic 1990). A classic example of Marr's approach is the paper by Marr and Nishihara in this volume. Marr borrowed from the work of the experimental psychologist J. J. Gibson and emphasized the importance of careful analyses of the visual input to understand the "problems" that must be solved by a vision system with abilities like ours. Marr went a step beyond his predecessors by showing how such observations could be used to analyze what computations are necessary to produce specific behaviors. And Marr did even more than this: His analyses were informed by facts about neurophysiology and neuroanatomy.

Marr stressed the idea that neural computation can be understood at multiple levels of analysis. Philosophers of science observed long ago that a single phenomenon can be examined at multiple levels of analysis (e.g., Putnam 1973; see also chapter 8 of Churchland 1986). When considering psychology, philosophers such as Jerry Fodor (1968) distinguished between a functional and physical level; the functional level ascribed roles and purposes to brain events, and the physical level characterized the electrical and chemical characteristics of those events.

Marr took these earlier analyses several steps further. He posited a hierarchy of levels that is rooted in the idea that the brain computes. He divided the functional level into two levels, one that characterizes *what* is computed and another that characterizes *how* the computation is accomplished (i.e., the algorithm), and he showed how these levels related to the lowest one, the level of the implementation. The theories of the computation and the algorithm are cast in the vocabulary of computation, referring to data structures and processes that operate on them, whereas the theory of the implementation is cast in the vocabularies of biology and biophysics, referring to properties of neurons and their physical interactions.

Marr's analysis held enormous appeal. Part of this appeal came from the idea that we could understand cognitive function by reason alone. If we could divine the correct theory of what had to be computed to solve a specific problem, we were most of the way there. Unfortunately, this has not turned out to be the case. The distinction between the levels of the algorithm and the implementation is not clear-cut. Indeed, if one is trying to understand brain function, then theories of computations and algorithms are necessarily descriptions of neural activity. Hence, properties of neurons cannot help but bear on these theories. It is clear that the structure of the brain (its neuroanatomy) and its dynamic properties (its neurophysiology) provide hints as to what the brain is doing. Marr himself noted the receptive field properties of different types of neurons and used this information to constrain his theory.

The interdependence of the levels has been driven home by the rise of "neural network" modeling since Marr's death (e.g., Grossberg 1987; Rumelhart and McClelland 1986). These models conflate all three levels. The functional properties of these models, several of which are described in articles we have reprinted here, are intended to mirror those of the underlying neural substrate.

Thus it became clear to many that research in brain function required a combination of complementary approaches, and the idea of a new field of research was born. The term "Cognitive Neuroscience" was coined in 1970 by Michael S. Gazzaniga, presently at the University of California, Davis. This field has since evolved as new technologies became available and the field became more interdisciplinary (Kosslyn and Koenig 1992), as we discuss in the following section.

The Cognitive Neuroscience Approach

The three foundations of cognitive neuroscience—neuroscience, experimental psychology, and computer science—have melded into a new discipline. The cognitive neuroscience approach can be schematized by a triangle. At the top of the triangle is behavior, and at its lower vertices are neuroscience and computation.

Consider first the top vertex. Our goal is to understand the regularities in how a system (vision, memory, etc.) behaves. Thus the behavioral sciences play a critical role in cognitive neuroscience; cognitive psychology, linguistics, psychophysics, and related disciplines provide detailed descriptions of what the brain does. Not all of the information produced by researchers in these fields is equally useful for cognitive neuroscientists, however; because our goal is to understand how the brain produces behavior, we need descriptions of behavior that can be related relatively directly to underlying neural mechanisms. Many sorts of behavior reflect complex interactions among many mechanisms. Unfortunately, there is no way to know in advance which regularities in behavior will mesh neatly with distinct properties of the brain and which regularities will reflect complex interactions among numerous properties. In general, however, behavioral studies that are designed to bear on issues about the brain seem more likely to alert us to relevant behavioral phenomena than studies that ignore such issues.

In cognitive neuroscience, explanations of behavioral regularities hinge on a confluence of facts and concepts about the brain and computation, which we schematize at the lower two vertices of the triangle. Our goal is to understand how a machine with the physical properties of the brain can produce specific behaviors when given specific inputs. Neuroscience, then, is the second vertex of our triangle; it provides information about the neuroanatomy and neurophysiology of the brain. Anatomical facts are often critically important because they specify the information flow within the brain, which places strong constraints on what a given part of the brain can do. And neurophysiological findings provide hints about how the brain represents and processes information.

Neuroscience has undergone a split; the part of this field that is concerned with behavior is now closely allied with cognitive neuroscience, and the other part has focused on the brain for its own sake. Even this second branch of the field, however, is beginning to feed into cognitive neuroscience. For example, the powerful methods of molecular biology are now being applied in studies of the neurochemical bases of brain function. In the future, we will understand brain function at multiple levels of analysis, ranging from the activity of circuits to the activity of genes that produce the neurotransmitters that are essential in making circuits work. The section on memory in this volume provides good examples of how cognitive functions are being approached from many levels. Memory is beginning to be understood from the levels of the transmitter, synapses and receptors (Bliss and Lømo), the function of neural circuits (Ambros-Ingerson, Granger, and Lynch; Barto and Jordan; Funahashi, Bruce, and Goldman-Rakic; Fuster and Jervey; Gluck and Thompson; Gnadt and Andersen; Hawkins and Kandel; Miyashita and Chang), and the behavior of the whole organism (Corkin; Mishkin; Schacter; Shimamura et al.; Squire).

Computer science is the third vertex of our triangle; it has given us not only the idea of computational analyses but also the possibility of computer models. As noted earlier, computational analyses lead to a theory of how input can be converted to output by a mechanism; such analyses are based on careful considerations about what kinds of input/output transformation would be necessary to produce a specific kind of behavior. However, in cognitive neuroscience we are not interested in any physical system that could perform computations and produce behavior. Rather, we are interested in one particular system, namely, the brain. Hence, computational analyses must be informed by neuroanatomy and neurophysiology.

The brain is very complex, as are the ways in which even relatively simple organisms can behave. Thus after computational analyses have led us to hypothesize specific computations, we often need *computer models* to discover the implications of these hypotheses. A computer model is a program that is designed to mimic a dynamic system. The program plays the same role as that played by a model aircraft in a wind tunnel when a new airplane is being designed. By observing the way the model behaves, researchers generate accounts for the way the actual object or system behaves and generate predictions about how it should behave in novel circumstances.

In cognitive neuroscience, models of neural networks allow researchers to discover how specific types of inputs can be mapped to specific types of outputs. The recent advent of neural network computer models has led to an explosion in modeling complex neural circuits; some of these models can produce remarkably brainlike behaviors and sometimes can offer insight into the underlying processes that produce these behaviors. Indeed, as is illustrated in several of the papers in this volume (Kosslyn et al.; Lehky and Sejnowski; Zipser and Andersen), the network models themselves can be analyzed after they "learn" to perform the mapping, providing the researcher with insights into what aspects of the input were used to achieve the mapping. These network models have also helped us to understand how individual nerve cells can show a very broad selectivity but the network as a whole can be very precise (see Kosslyn et al., this volume).

Finally, we must note that the invention of more powerful computers has in turn engendered advances in medical technologies, which have had a large impact on the study of human brain function. Thanks to CT and MRI scans, clinicians now know exactly where brain damage is located when they evaluate cognitive functions and deficits of their patients following brain injury. Moreover, new brain scanning techniques allow researchers to observe which specific regions of the brain are active while human subjects perform specific tasks. Some of these methods (such as positron emission tomography (PET) and magnetoencephalography (MEG)) are illustrated in articles we have reprinted in this book (Kaufman and Williamson; Petersen et al.; Roland and Friberg).

We have schematized the interactions among these fields with a triangle because each pair of considerations may cross-fertilize. For example, discoveries about behavior can lead researchers to make discoveries about the brain; for instance, the existence of "illusory contours" (visible edges that do not actually exist) allowed researchers to study where in the brain such contours are provided (von der Heydt et al., this volume). And such discoveries can provide direct hints as to how to build a computational system that behaves as we do—one goal of AI. Similarly, the concepts and facts schematized at both bottom vertices can lead researchers to discover more about how the system behaves; for example Lowe's (1987a, 1987b) computational ideas about visual encoding led Biederman (1987) to perform a series of experiments on the properties of human perception. The finding that the visual system constructs surfaces from sparse data and that it deals with motion transparency through early segmentation of surfaces led to new computer algorithms for analyzing structure-from-motion that are much more powerful than previous algorithms (Andersen et al. 1991).

In short, cognitive neuroscience involves an interplay between three different kinds of concepts and findings. No discipline or approach is paramount; they all lean on each other. Furthermore, no single type of theory or analysis necessarily precedes the others; we need not begin with a particular behavior in mind in order to discover something interesting about the brain that ultimately helps to explain behavior. The field necessarily involves a dynamic interplay between three kinds of information.

Although it is difficult to categorize all scientific investigations in a field, for purposes of clarity we will attempt to distinguish cognitive neuroscience from computational neuroscience, cognitive science, and some areas of research within the field of neuropsychology. Of course, there is a great deal of overlap between these fields and cognitive neuroscience. First, cognitive neuroscience and *computational neuroscience* differ in the kinds of questions that are asked and the kinds of answers that are sought. Computational neuroscience focuses on specific problems, such as stereopsis or motion detection, from a relatively abstract perspective and typically seeks an answer at the level of what Marr called the algorithm. The approach tends to focus more on the mathematical solution of a problem than on the detailed brain mechanisms or behavioral correlates involved.

Second, cognitive neuroscience also differs from *cognitive science* in its style of investigation. Cognitive neuroscientists ask questions about how the brain produces behavior; that is why "neuroscience" is the noun. In contrast, cognitive science focuses on function per se, with little specific regard for the brain. A cognitive neuroscientist attempts to characterize the functions of neurons or networks of neurons, whereas a cognitive scientist does not.

Third, cognitive neuroscience differs from some research avenues in contemporary *neuropsychology*. Contemporary neuropsychology has split into three major groups. (1) Clinical neuropsychologists are interested primarily in diagnosing the effects of brain damage so that effective rehabilitation programs can be designed. To the extent that these clinicians perform research, it tends to be descriptive (documenting patterns of deficits) and often does not rely on rigorous formal experimentation or theorizing. (2) Another part of neuropsychology evolved into *cognitive neuropsychology* (which is exemplified by articles published in the journal of the same name). These researchers attempt to characterize mental function in its own right, typically by studying the effects of brain damage on behavior; however, they are not as interested in characterizing what is contributed by specific parts of the brain. Thus cognitive neuropsychologists ask the questions similar to those of cognitive scientists, with little concern about the exact implementation of functions by neural structures. Cognitive neuropsychologists differ from cognitive scientists in two ways: they try to characterize mental function by observing selective behavioral deficits following brain damage, and they rely less heavily on computational approaches. (3) Finally, the third wing of neuropsychology represents a major force in the field of cognitive neuroscience. These researchers have adopted neuroanatomical, neurophysiological, and computational perspectives in order to understand how the brain gives rise to mental activity. Several examples of this approach are the articles by Coltheart and colleagues, Corkin, Farah, Gazzaniga and colleagues, Geschwind, Levine, Marshall and Newcombe, McCarthy and Warrington, Milner and Petrides, Mishkin, Schacter, Shimamura and colleagues, and Squire, in this volume.

Structure of the Book

Cognitive neuroscience is inherently interdisciplinary. It draws on concepts, findings, and methods from fields that focus on behavior, neuroscience, and computation; the brain is regarded as a living machine that produces behavior. One of the exciting aspects of this field is that anyone entering it will find something new. But this variety is also one of its drawbacks. It is often difficult to come into an area and discover the relevant literature, let alone determine which papers are of most importance. The purpose of this book is to make it easier for researchers and students to enter cognitive neuroscience by pulling together many of the key articles that form the foundations of the field.

A survey of the papers included in this volume will show that the field of cognitive neuroscience continues to approach many of the traditional questions in psychology and neuroscience, reshaping and refining them as the field evolves. The book begins with key articles in the cognitive neuroscience of vision. Because vision has been such an active area, we divided this part into two sections, one that focuses on the nature of distinct "streams" of visual processing and one that focuses on the ways in which neurons code information. We next present articles on other sensory modalities, specifically audition and somatosensory processing. We then present a set of articles on attention. Research on this higher cognitive function has enjoyed success in part because of the ability of researchers to define the problem and develop inventive behavioral tasks. We then turn to memory; this part is also relatively long, and it has been divided into a section on the mechanisms used to store new information and another section on the nature of distinct memory systems. We close with a part on higher cortical functions, which we organize into a section on reasoning and a section on language.

Although the field is very young, we were surprised by the large number of first-rate articles that have already been published. Thus we had some difficult choices to make, given the limited space available. Some of our choices were influenced by our observation that the advent of neural network computer models has sometimes had an unfortunate consequence: Some researchers have tended to focus on the requirements of simply building a system that produces a behavior and have not exploited hints from the brain or heeded its warnings. Mindful of this shortcoming, we have focused in this volume on reprinting papers that bear directly on how the brain functions. "Neuroscience" is, after all, the noun. We have chosen papers that examine specific aspects of behavior, either explicitly or implicitly, and that cast their findings in a way that provides direct constraints on computational theories. A minority of papers actually present computational models, in part because such models have only recently been brought to bear on many issues in the field. But all of the papers in this volume provide the foundations for further research in cognitive neuroscience, and it is our goal that this volume will allow more researchers to join us in the enterprise.

We hope that the reader will come away with an appreciation not only of the complexities but also of the excitement that is generated when a new field evolves—one that is ready and willing to incorporate technical advances and use them to attack one of our final frontiers.

References

Allman, J. M., J. F. Baker, W. T. Newsome, and S. E. Petersen. 1981. Visual topography and function: Cortical visual areas in the owl monkey. In *Cortical Sensory Organization*. Vol. 2, *Multiple Visual Areas*, ed. C. N. Woolsey, 171–186. Clifton, NJ: Humana Press.

Andersen, R. A., R. J. Snowden, S. Treue, and M. Graziano. 1991. Hierarchical processing of motion in the visual cortex of monkey. *Cold Spring Harbor Symposia on Quantitative Biology* 55: 741–748.

Anderson, J. R. 1978. Arguments concerning representations for mental imagery. *Psychological Review* 85: 249–277.

Biederman, I. 1987. Recognition-by-components: A theory of human image understanding. *Psychological Review* 94: 115–147.

Bolz, J., C. D. Gilbert, and T. N. Wiesel. 1989. Pharmacological analysis of cortical circuitry. *Trends in Neurosciences* 12: 292–296.

Broca, P. 1863. Localisations des fonctions cérébrales.—Siège du langage articule. *Bulletins de la sociéte d'anthropologie* 4: 200–204.

Brodmann, K. 1909. *Vergleichende Lokalisationlehre der Grosshirnrinde in Ihren Prinzipien Dargestellt auf Grund des Zellenbaues.* Leipzig: A. J. Barth.

Churchland, P. S. 1986. *Neurophilosophy: Toward a Unified Science of the Mind-Brain.* Cambridge, MA: MIT Press.

Flourens, P. 1824. *Recherches expérimentales sur les propriétés et les fonctions du système nerveux dans les animaux vertébrès.* Paris: Crevot.

Fodor, J. A. 1968. *Psychological Explanation: An Introduction to the Philosphy of Psychology.* New York: Random House.

Fritsch, G. T., and G. Hitzig. 1870. Uber die elektrische Erregbarkeit des Grosshirns. *Archiv für Anatomie, Physiologie und Wissenschaftliche Medizin* 37: 300–332.

Gall, F. J. 1812. *Anatomie et physiologie du système nerveux en général, et du cerveau en particulier.* Paris: Schoell.

Grossberg, S., ed. 1987. *The Adaptive Brain.* New York: North-Holland.

Kosslyn, S. M., and O. Koenig. 1992. *Wet Mind: The New Cognitive Neuroscience.* New York: Free Press.

Kosslyn, S. M., and V. Maljkovic. 1990. Marr's metatheory revisited. *Concepts in Neuroscience* 1: 239–251.

Lashley, K. S. 1929. *Brain Mechanisms and Intelligence.* Chicago: University of Chicago Press.

Lashley, K. S., and G. Clark. 1946. The cytoarchitecture of the cerebral cortex of Ateles. *Journal of Comparative Neurology* 82: 233–306.

Lowe, D. G. 1987a. Three-dimensional object recognition from single two-dimensional images. *Artificial Intelligence* 31: 355–395.

Lowe, D. G. 1987b. The viewpoint consistency constraint. *International Journal of Computer Vision* 1: 57–72.

Luria, A. R. 1973. *The Working Brain.* London: Allen Lane.

Luria, A. R. 1980. *Higher Cortical Functions in Man.* New York: Basic Books.

McCulloch, W. S., and W. Pitts. 1943. A logical calculus of the ideas imminent in nervous activity. *Bulletin of Mathematical Biophysics* 5: 115–133.

Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* San Francisco: W. H. Freeman.

Marr, D., and H. K. Nishihara. 1978. Visual information processing: Artificial intelligence and the sensorium of sight. *Technology Review* 81: 2–23.

Marr, D., and T. Poggio. 1976. Cooperative computation of stereo disparity. *Science* 194: 283–287.

Nottebohm, F. 1981. Laterality, seasons and space governing the learning of a motor skill. *Trends in Neurosciences* 4/5: 104–106.

Penfield, W., and P. Perot. 1963. The brain's record of auditory and visual experience. *Brain* 86: 595–697.

Posner, M. I. 1978. *Chronometric Explorations of Mind.* Hillsdale, NJ: Erlbaum.

Putnam, H. 1973. Reductionism and the nature of psychology. *Cognition* 2: 131–146.