Introduction:
Four Themes in Research on the Neurobiology of Memory

STUDY QUESTIONS

What is the neurobiology of memory?

What major questions about memory are pursued with a neurobiological approach, and how are these questions addressed in experimental analyses?

What are meant by "connection," "cognition," "compartamentalization," and "consolidation"?

Our memories reflect the accumulation of a lifetime of experience and, in this sense, our memories are who we are. Surely the background of our makeup is determined largely by our genes; genetics sets the range of what we can aspire to be. However, by contrast to generality of genetic limitations, the specifics are a matter of memory. We learn to walk, to dance, to drive a car, to throw a ball, and to play a video game—a myriad of acquired skills we come to take for granted. We learn to fear dangerous situations, to appreciate particular types of music and styles of art—a broad range of aversions and enjoyments we have assumed as elements of our preferences and personality. We learn to speak, and to speak and understand our particular language. We learn world history, and we learn our own family tree and personal autobiography—all of these, and much, much more, compose the vast contents and intricate, complex organization of memories that make each of us a unique human being. So, the
Introduction

analysis of memory is a search for self-understanding, an adventure that promises to reveal the inner secrets of how we came to be who we are.

The nature of memory—its basic biological structure, its psychological character and organization, and its longevity—has been the subject of investigations by philosophers, writers, and scientists for hundreds of years, and each approach offers its own distinct avenue for understanding memory. Recently, with the rise of modern methods of cognitive science and neuroscience, and their combination, many new and deep insights about the mechanisms of memory have emerged. These observations have also led us to a greater general understanding about the mind and about brain functions that mediate cognition, emotion, behavior, and consciousness. The aim of this book is to explore memory from the perspective of cognitive neuroscience, offering a historical and a current overview of how brain functions in memory have been studied and what we have learned about memory as an encompassing aspect of the mind.

The present chapter introduces some of the philosophical and historical underpinnings of research on the biological bases of memory. I begin by presenting four central themes that have guided memory research for over a hundred years. Substantial preliminary evidence regarding each of these themes emerged during a “Golden Era” for neuroscience in the latter half of the nineteenth century and the beginning of the twentieth century. A brief introduction to some of these accomplishments provides the background for a subsequent, more detailed summary of progress on each of the four central themes in the remainder of the book.

The four “C’s”

At the outset of systematic investigations of the nervous system four major themes dominated considerations of brain function of relevance to memory. In part to facilitate your memory for them, I refer to these themes as the four “C’s”: connection, cognition, compartmentalization, and consolidation.

The first theme—connection—concerns the most basic level of analysis of memory function, the basic nature of the circuitry of the brain including the elements of information processing and how they communicate with one another in the service of memory. The emphasis on “connection” here reflects the major conclusion that has emerged in this research—that memory is encoded within the dynamics, that is, the changeability or plasticity, of connections between nerve cells. More specifically, the consensual view from the perspective of memory as a phenomenon of brain cells is that memories are instantiated by alterations of the strength or reliability of communication between cells via their synaptic connections. Efforts to understand the nature of the storage mechanism are aimed at identifying the biological materials, the cellular “switch,” from which memories are made. Accordingly, cellular and molecular analyses of memory have been focused on characterizing the modifiability or plasticity of neural connections that underlie memory performance. Experimental research in this area seeks to discover specific molecular events that form the basis of a memory, and subcellular structures that support neural connections that accomplish memory. This research also asks how the basic memory storage mechanism can be modulated by natural neurochemical events and manipulated by genetic and pharmacological interventions.

The second theme—cognition—refers to the nature of memories at the highest level of analysis, the psychological level. Consider the following unusual instance of memory. You walk into a new room, an odd colored light fills the room, then a loud and frightening rattling noise goes off and persists—you exit with haste. Subsequently, you happen to enter that room again, and again the odd light comes on. You immediately leave. Is the memory for that event represented in terms of a new association between particular novel sensory stimulus, the odd light, and the escape behavior you executed in response to it, what psychologists call a “stimulus–response association”? Or is the fundamental association between the light and the consequent irritation the loud sound evoked in you, what psychologists call a “stimulus–reinforcer association”? Or are the concepts of stimulus–response and stimulus–reinforcer associations altogether too simplistic a view of how your memory is stored? Does your memory representation contain a record of the entire series of relevant and peripheral events that constituted the learning episode? Is that memory isolated among a large and loose collection of episodic memories? Or is that memory contained within a systematic organization of other experiences in the same building that form a network of knowledge about your experiences there?

Questions about the nature of memory representations have been at the center of debates about whether the complexity of memory can or cannot be reduced to a set of simple associative principles. During the Golden Era there were divergent and strong views espousing either that memory can be simplified to a set of principles about stimulus–response and stimulus–reinforcer associations or that memory involves complex networks that can be understood only in terms of cognitive operations that are not reducible to simple associative mechanisms. The understanding that has emerged from this area of research is that there are mechanisms that guide behavior from stimulus–response and from stimulus–reinforcer associations, and there is a “cognitive” form of memory that is distinguished in
both its psychological mechanisms and anatomical pathways from the other forms of memory.

The third theme—compartmentalization—addresses the question of memory localization. This question appeared early on in debates about whether memory can be localized to a particular area of the brain, especially the cerebral cortex, or whether it is distributed throughout the cortex, or indeed the entire brain. The major conclusion that has emerged on this theme is that memory as a whole is distributed widely in the brain. But, at the same time, different kinds of memory are accomplished by specific brain modules, circuits, pathways, or systems. That is, memory is compartmentalized. The compartmentalization occurs at two levels. First, the cerebral cortex is composed of many anatomically circumscribed “modules,” each of which makes a correspondingly specific contribution to memory function. Second, there are multiple memory systems in the brain, all of which involve the cerebral cortex but they diverge in pathways leading from the cortex to other structures that lie beneath the cortex (subcortical structures), and the systems formed by these cortical–subcortical pathways accomplish different kinds of memory. This research area also considers the relationship between memory and other cognitive processes, including consciousness, coordinated movement, and emotion. The major aim of research on brain modules and systems is to identify and distinguish the different roles of specific brain structures and pathways, usually by contrasting the effects of selective damage to specific brain areas. Another major strategy in attacking this issue focuses on localizing brain areas that are activated, that is, whose neurons are “turned on” during particular aspects of memory processing. Some of these studies use new functional imaging techniques to view activation of brain areas in humans performing memory tests. Another approach seeks to characterize the “code” for memory within the activity patterns of single nerve cells in animals, by asking how information is represented by the activity patterns within the circuits of different structures in the relevant brain systems.

The fourth theme—consolidation—concerns when and how memories become permanent. It is well known that some experiences are rapidly forgotten, whereas others are remembered for a lifetime. And there have been many anecdotal and clinical reports that various forms of interference or head or brain injury can “wipe out” memories that were recently acquired but have less effect on memories acquired remotely before the interfering event or injury. These observations suggest that memories are initially labile and later become resistant to loss, suggesting a process of consolidation during which memories take on a permanent form. Modern research has shown that there are two general kinds of consolidation. One of these—which I call “fixation”—involves a cascade of molecular and cellular events during which the changes in connections between cells become permanent in several minutes to hours after a memory is formed. This process can be influenced by many factors, including among them a specific brain system for modulation of memory fixation. The other kind of consolidation process is called “reorganization” because this process involves a prolonged period during which distinct brain structures interact with one another, and the outcome is that newly acquired information is integrated into one’s previously existing body of knowledge. This reorganizational process therefore involves an entire brain system, and discovering how it works involves a consideration of both the individual contributions of particular parts of the brain system and the nature of interactions among the parts.

The Golden Era

The aim of this chapter is to set the stage for the succeeding sections that will review our understanding of each of the four major themes in memory research, and in doing so provide a framework for understanding the neurobiological bases of memory. I pursue an historical approach, elaborating on each of the four “Cs,” beginning with a summary of discoveries made at the threshold of modern neuroscience research on memory in the latter half of the 1800s.

Before that period, some of the critical background had already long been established. In 1664 the early anatomist Willis published the first description of the anatomy of the brain and had suggested that different brain areas controlled distinct functions. Simplified views of the brain and its main components are provided in Figures 1–1 and 1–2. Also, in 1791 Galvani introduced the notion that electricity is the mechanism of nervous conduction. But around the turn of the twentieth century several additional major discoveries formed the full beginning of a scientific analysis of the brain, a Golden Era in which several key findings led toward real progress in understanding brain function and memory. Some of these contributions represented major advances to the field of neuroscience in general, and others pertained to memory research in particular. Here I highlight a few of the major insights of that period that have had lasting impact. Put together, these observations should give the reader a sense of the field of neuroscience, and especially about the neurobiology of memory, upon which all subsequent progress is based.
Connection: Cellular substrates of brain communication and memory

One main set of advances that laid the foundation for memory research involved discoveries about the basic building blocks of brain circuits. These discoveries identified the fundamental elements of the brain, characterized how they are assembled into simple functional circuits, and demonstrated that they can be modified during learning.

The neuron doctrine

One major area of discovery about the brain that contributed directly to our understanding of the cellular and molecular substrates of memory was the development of the "neuron doctrine" by the Spanish anatomist Santiago Ramon y Cajal. The neuron doctrine is the notion that the brain is composed of discrete nerve cells, and that these cells are the essential units of information processing, connected to one another so as to transfer and integrate information in large-scale networks. At the time of Cajal, this view was not entirely new, but it was also not widely accepted. Rather, Cajal's work addressed a major controversy about the nature of connections between neurons. In the debate, one camp, called the "reticularists," argued that the brain is a single unified interconnected network of fibers in which all the cells were fused to one another. By contrast, the other camp, called the "antireticularists," suspected that the brain was composed of independent nerve cells as units, but they had no definitive evidence. Before Cajal, the strongest argument for independent cells came from the observation that small lesions in one area resulted in sharply defined areas of degeneration, not what one would expect of a fused network.

Cajal's success was based on his adoption of a new staining method that was developed by another anatomist named Camillo Golgi in 1873. The method involved a "black reaction," a new silver stain that had the remarkable quality of darkening the entire cell membrane of a neuron. At the same time, the staining was selective to only a small fraction of the neuron population in an area of brain tissue. Thick sections of the brain stained this way provided a full view of individual cells standing out clearly against the background of many other surrounding pale cells. Using this method Cajal was able to provide the most striking confirmation of the already existing identification of the major elements of nerve cells. As shown in Figure 1–3, these include the cell body, the multiple fine processes that extended from one end of the cell body called dendrites, and the single larger process extending from the other end of the cell body called the axon. Also, he noted the specialization of the axon as it contacted the dendrites of other cells; this specialization would later be called the synapse (see next section).

Cajal attempted many variations of the procedure. He found that tissue with less myelin, the insulation layer of axons, produced the clearest images of neuronal processes, and the best cases were found in young brains and in birds. He also found that thicker sections allowed one to examine all of the extensions of the cell membrane that connect with other
cells. In each of several different preparations he found elaborate endings of axons in nests or baskets of axonal "arborization"—a treelike branching—of the axon connecting it to multiple parts of another cell or multiple cells. In no case did he observe the stain continuing into the next cell, as would be expected if there was a fusion of the axonal ending with dendrites of another cell. Cajal concluded that there must be some method of communication between cells that did not involve a joining of their membranes. Cajal's preparations and the evidence they provided were elegant, and convinced other anatomists and physiologists that each cell was contained within a membrane and was separate although in contact with other cells. These observations won him the Nobel Prize in 1906.

Cajal was also able to make key conclusions about the function of neurons from his observations on their anatomy. As said previously, he confirmed the existence of all of the major components of the nerve cell. Moreover, in his studies on visual and olfactory sensory structures, Cajal noted that the dendrites pointed toward the outside world and that the axon pointed toward the brain. From these observations he deduced that nerve cells were functionally polarized, such that information flows from the dendrites to the axon and is subsequently conveyed by the specialized connection to the dendrites of another nerve cell.

These conclusions established the essential view that the integration of information occurred by the summation of signals converging from the axons of several neurons onto the dendrites of cells receiving those inputs. In addition, Cajal developed some important and prescient ideas directly relevant to the basic memory mechanism. He studied the brains of several species and observed that the vertebrates higher in the phylogenetic scale had a greater number of connections between nerve cells. He concluded that the increase in connectivity could be the basis of greater intellectual power of the higher species. He suggested that mental exercise could facilitate increased connectivity through a greater number and intensity of the connections, and that these changes in connectivity could coincide with the acquisition of skills such as playing a musical instrument. Cajal's insights have become key axioms for the study of neuronal function and communication. Moreover, his suggestion that "plasticity" in number and strength of neuronal connections underlies learning guides the search for molecular and cellular substrates of memory.

The reflex arc

In the same period other major advances were made from studies on the physiology of the nervous system. Perhaps most important among these were Charles Sherrington's observations on the nature of reflexes in the spinal cord. Before Sherrington the existence of involuntary muscle actions was already well recognized, including the basic observation that specific sensory stimulation could be "reflected," as if by a mirror, to generate muscle movements—hence the "reflex arc" (Fig. 1-4). In addition it was known that complex reflexes, such as those mediating jumping in frogs or coordinated flying movements in birds, could be elicited even following decapitation, suggesting control of complex coordination could happen at a level below the brain—at the level of the spinal cord. And it was clear that reflex arcs accomplished within the spinal cord could be inhibited by higher level control. However, there was very little understanding of the underlying circuitry that accomplished either simple or complex reflexes.

Sherrington made many contributions that provided the foundations for our understanding of neural circuitry that are as relevant today as they were when he made his discoveries. Even before Cajal's convincing anatomical evidence was provided, Sherrington had reached the conclusion that neurons must be independent elements. Part of the evidence came from his studies on neural degeneration, showing that cortical lesions, damage induced by heat or cutting, resulted in restricted, not diffuse patterns of degeneration. He also realized that the neuron doctrine, and the detailed evidence showing the connection was from axons to dendrites, could explain why neural transmission was one-way. And the discontinuity between cells provided a mechanism for why there was a time lag in reflexes such that they were much slower than predicted from the speed of conduction of the neural impulse—the loss of time in long-range conduction had to involve the extra time re-
Introduction

Figure 1-4. A schematic diagram of some components of the pain withdrawal reflex arc. Pain to the skin activates a sensory neuron whose cell body is just outside the spinal cord. The axon synapses with an interneuron within the spinal cord, which excites a motor neuron that causes contraction of the relevant flexor muscle, and excites an inhibitory interneuron that decreases activity in motor neurons that activate the extensor muscles in the same limb.

required for an impulse to jump a “gap” between the neural elements. Sherrington is credited for inventing the term *synapse* to describe the hypothetical junction between neurons where transmission occurs.

In his classic studies on the “knee jerk” reflex, Sherrington demonstrated the details of his model reflex arc, wherein specific sensory information is gathered at the input end of the arc and then relayed to turn specific muscles on and off at the output end. He provided critical evidence for the existence of the “sixth sense”—specialized sensory receptors in the muscle that monitor muscle length and tension, and showed that these sensory elements send nerves into the spinal cord. He also contributed to the characterization of the maps of sensory inputs from the skin surface into the spinal cord, the so-called dermatomes. By isolating and stimulating sensory roots he could determine which regions of the skin evoked reflexive movements. Conversely, to map the output pattern, he stimulated individual spinal cord motor roots and characterized which muscles were activated.

Important as these discoveries were, perhaps Sherrington’s most outstanding contribution was elucidating some of the complexities of the circuitry that underlie reflexes. He concluded that, even at the level of the spinal cord, reflex arcs are not entirely separate or independent. In these circuits the initial sensory neurons in the spinal cord receive dedicated inputs, but the output neurons of even the simplest reflexes receive information from many of those input cells (Fig. 1-4). Thus, consistent with Cajal’s observations on the anatomy of nerve cells, there was a summation or integration of both inputs into a common path that could guide complex behavioral actions. And there were two major kinds of inputs that combined in this integration. There were excitatory inputs by which an impulse from the input cell increased the likelihood of impulse generation in the next cell. And there were inhibitory inputs by which an impulse in the input cell decreased the likelihood of impulse generation in the next cell. He also made major discoveries about the phenomenon of reciprocal innervation, the basic mechanism by which excitatory and inhibitory influences are coordinated to mediate movement. Within this scheme every action by one muscle is coordinated with an opposing action of complementary muscles. For example, the simple act of walking involves the coordinated actions of flexing one group of muscles during the extension of others, followed by the opposite complementary actions in executing the next step in walking.

In addition, Sherrington employed the technique of decerebration, cutting the spinal cord just behind the brain, to show that complex coordinated actions existed even without higher level cerebral control. He showed, for example, in a decerebrate cat, when a forelimb was excited to move forward, the hindlimb on the same side moved back and the two legs on the other side of the body exhibited the opposing movements, producing the pattern seen in normal walking, but without conscious cerebral control. In studies on the “scratch reflex” he showed the specificity of arm movements for scratching evoked in response to small areas of skin stimulation. Furthermore, in extensions of these studies he revealed that coordinated and directed scratching movement patterns could be played out over time, with alternating extension and flexion of muscles at different levels of the limb to produce repeated scratching movements, plus postural adjustments in the other limbs to support standing without the use of the scratching limb.

Sherrington envisioned all of this as accomplished within the spinal cord, by a “chaining” of reflexes wherein successive coordinated movements are elicited by their predecessors. These elements provided the outline for his formulation on the integrative action of the nervous system. In this prototype for modern views, Sherrington proposed a hierarchy of coordinated control wherein the cerebral cortex is acknowledged as the
newest and most complicated switchboard of reflexes, and successively lower centers in the brain stem and spinal cord to mediate more and more specific, yet still complex and coordinated actions.

**The conditioned reflex**

Initially unrelated to this course of research, the Russian physiologist Ivan Pavlov was making landmark discoveries about the nervous control of digestion. He discovered that the release of digestive fluids was controlled by the nervous system, contrary to the prevailing view that digestive fluids were released as a consequence of mechanical stimulation by food directly onto the stomach wall. To test his hypothesis that the nervous system was involved, Pavlov developed a surgical procedure in which he severed the gullet and attached both open ends to the skin of the neck. This allowed him to either introduce food into the mouth and upper gullet and then retrieve it without going to the stomach, or introduce food directly into the stomach. He found that food stimulation associated with ingestion caused the release of gastric fluids in the stomach, even when the food never reached its normal target. He concluded that food excites the gustatory sensory apparatus in the mouth and gullet, transmitting signals into the brain stem, which, via the vagus nerve, controls the release of gastric fluids. From a neurophysiological perspective, one can say that Pavlov identified a reflex arc for digestion, and for this he received the Nobel Prize in 1904.

But by the time he received the prize, Pavlov had already turned his interest to an intriguing report that gastric juices of a horse could begin to flow not only with a direct application of gustatory stimuli but also even when the animal only caught sight of hay. Pavlov replicated the phenomenon of “psychic secretion” using dogs and measured the generation of fluids from the salivary glands. He found that the sight of a piece of beef indeed caused salivation, but he also found that the phenomenon was unreliable. The salivation tended to decrease following repeated presentations of the sight of beef. Conversely, sometimes salivation was initiated by events that preceded the sight of beef, for example, when the person who regularly provided the food merely appeared in the testing room.

Pavlov set out to meticulously control the stimuli available to the animal, and he tried out many arbitrary stimuli—including the famous bell rung prior to the presentation of food. Based on the results from a broad range of experimental manipulations, Pavlov concluded there were two kinds of reflexes. One kind of reflex is “unconditioned,” identical with the innate and stable reflexes of Sherrington. The unconditioned reflex is composed of a particular unconditioned stimulus (US) that inevitably elic-

its its characteristic unconditioned response (UR). The other kind of reflex is “conditioned,” that is, acquired through experience. This kind of reflex is the unstable one, and is composed of an arbitrary conditioned stimulus (CS) that when paired with a US comes to elicit a conditioned response (CR) similar in form to the UR. Pavlov identified the critical importance of two parameters in establishing and maintaining the conditioned reflex: the close temporal contiguity of the CR and US—the CS must lead the US by a particular time interval, and the CS must consistently predict the US.

Combined, the contributions of Cajal, Sherrington, and Pavlov, as well as many others, laid the basic framework for succeeding views of brain circuitry and function, as well as its role in memory. They showed that the basic elements of the circuits are independent neurons that communicate across synapses, that these elements are integrated within complex patterns of circuitry for coordinated action built up from simple reflex arcs, and that these circuits can be modified to support learned reflexes. These insights set the stage for future investigations on the mechanisms of how connections between neurons are modified during learning, and guided the development of views on the organization of memory for both simple and complex behaviors.

**Cognition and memory**

During the same period, distinct developments were made toward characterizing the nature of memory from a purely psychological perspective. Two main and competing lines of theorizing developed. One school, called “behaviorism,” developed out of a desire to provide a rigorous science of memory consistent with the findings on the neurophysiology of conditioned reflexes, and attempted to explain all of learned behavior on the basis of elements of association and conditioned responses. The other school, called “cognitivism,” emphasized the complexity of learned behavior, and its promoters could not be persuaded that all aspects of cognition, insight, and planning could be captured in an elaborate account of associations or reflex chains and instead required a more elaborate conception and, correspondingly, a more complex neural instantiation.

**Behaviorism**

The tradition of rigorous methodology in memory research began with Herman Ebbinghaus, who admired the mathematical analyses that had been brought to the psychophysics of perception, and he sought to develop similarly precise and quantitative methods for the study of memory. Bas-
ing his work on a large number of pioneering studies, in 1885 Ebbinghaus published a monograph that set a new standard for the systematic study of memory. Ebbinghaus rejected the use of introspection as a methodology that was prominent in previous conceptual schemes about memory. In its place he developed several key new techniques that would control the nature of the material to be learned and provide quantitative objective assessments of memory performance. To create learning materials that were both simple and homogeneous in content Ebbinghaus invented the “nonsense syllable,” a meaningless letter string composed of two consonants with a vowel between (e.g., “ket,” “poc,” “baf”). With this invention he avoided the confounding influences of what he called “interest,” “beauty,” and other features that he felt might affect the memorability of real words. In addition, the nonsense syllable simultaneously equalized the length and meaningfulness of the items, albeit by minimizing the former and eliminating the latter. Furthermore, to measure memory Ebbinghaus invented the use of “savings” scores that measured retention in terms of the reduction in trials required to relearn material. In addition, he was the first to employ mathematical-statistical analyses to test the reliability of his findings.

It was also in this period that systematic studies on animal learning and memory had their beginnings. In 1901 Small introduced the maze to studies of animal learning, inspired by the famous garden maze at Hampton Court in London (Fig. 1–5). He began what would become an industry of systematic and quantitative studies to identify the minute details of how rats acquired specific responses in repetitions of turns taken in the maze. But he observed that within a trial or two rats prefer a shortcut over the response route that had been reinforced on many previous trials, leading Small to conclude that future experiments should investigate the natural biological character of the animal if one is to be able to interpret the findings. These initial observations set forth a major controversy in the field of animal learning. Can learning be reduced to a set of arbitrary associations between external stimuli and behavioral responses, or must one consider issues such as cognition, insight, and motive?

At the turn of the century, Edward Thorndike had invented a “puzzle box” in which he observed cats learning to manipulate a door latch to allow escape from a holding chamber. Based on his observations he proposed the “law of effect,” which stated that rewards reinforced repetitions of the specific behaviors that preceded them. (This simple law would be reinvented and extensively elaborated by B.F. Skinner in the 1950s to explain all of learned behavior.) In the same period John Watson published his accounts on maze learning by rats. In one of his most famous experiments, Watson trained rats to run a maze and then searched for the underlying stimulus control by eliminating one sense after another. He found the rats could still run the maze with only the kinesthetic (muscle) sense remaining, leading him to conclude that the learning must be mediated by a chain of reflexes, consistent with the evidence from physiology. By 1913 he had accumulated sufficiently compelling evidence for the reductionist strategy that he wrote a “behaviorist manifesto,” formalizing the view that learning could be understood in terms of simple stimulus and response associations without resorting to considerations of vague concepts such as consciousness.

**Cognitivism**

William James captured prevalent views of the time on the origins of both the behaviorist and cognitivist perspectives in his classic *The Principles of Psychology*. Within this treatise, James considered reflex mechanisms and pathways as the essential building blocks of memory, and he called these the mechanisms for the formation of a “habit.” James viewed habits as built upon a very primitive mechanism that is common among biological
systems and due to an inherent plasticity of organic materials. Within the nervous system he viewed the mechanism of habit in terms of its known electrical activity, and suggested that electrical currents should more readily traverse paths previously taken. Thus, James felt that a simple habit was nothing more than the discharge of a well-worn reflex path, entirely consistent with the views of emerging behaviorism. Furthermore, James expanded on this notion, attributing great importance to habits as the building blocks of more complicated behavioral repertoires. He suggested that well-practiced behaviors and skills, including walking, writing, fencing, and singing, are mediated by concatenated discharges in connected reflex paths, organized to awaken each other in succession to mediate the serial production of learned movement sequences.

However, while acknowledging the importance of habits as the fundamental mechanism that underlies memory, James recognized real “memory” as something altogether different from habit. James argued that there were two forms of memory, differentiated by their timing and by their role. He suggested that initially there is a primary memory, what we today call short-term or working memory, a short-lived state where new information has achieved consciousness and belongs to our stream of thought. Primary memory also serves as the gateway by which material would enter secondary memory, what we now call long-term memory. James emphasized that secondary memory involves both the intellectual content of information we have learned and the additional consciousness of the experience during learning.

In addition to the feature of personal consciousness, the full characterization of memory was framed in terms of its structure as an elaborate network of associations. James argued that, while memory is based on the habit mechanism, it is vastly elaborated such that the formation of associations among habits supports the richness of our experience of a memory. Thus, the underlying foundation of recall involves a complex, yet systematic set of associations between any particular item and many other co-occurring items during one’s experiences.

It is of interest that James also offered speculations that touched on the biological basis of memory. He suggested that memory depends on two aspects of the habit mechanism. First, how good a memory is depends on the strength or persistence of the pathway—this aspect James suggested was innate. Second, he suggested that memory depends on the number of pathways through which an item is associated. He emphasized the latter as more malleable, and argued that the key to a good memory is to build diverse and multiple associations with one’s experiences, weaving information and experiences into systematic relations with each other. The capacity to search through one’s network associations was held to be the basis of conscious recollection, and could lead to creative use of memory to address new problems. Conversely, James admonished his students not to simply rehearse learned materials. This, he argued, could lead only to the concatenation of habit pathways that could only be expressed by repetition.

James never contrasted “habit” and “memory” as distinct forms of memory. A more direct recognition of two forms of memory can be attributed to the philosopher Henri Bergson, who in 1911 explicitly proposed that representations of the past survive under two distinct forms, one in the ability to facilitate repetition of specific actions, and the other in independent recollections. The suggestion that habits and memories might both have a common cellular basis, and at the same time exist as distinct forms of memory, would ultimately resolve the controversy between the behaviorist and cognitivist schools. In addition, the notion of different forms of memory would reappear in the solution to the controversy about the compartmentalization of memory discussed next.

Compartmentalization of cortical function and memory

In this same period as advances were made in characterizing reflex circuitry, and the controversy about the nature of memory was brewing, a separate battle was engaged over another critical puzzle about the brain and memory. This controversy focused on the organization of the cerebral cortex. There was already scattered evidence from studies of patients with circumscribed brain damage that the anterior (front) and posterior (back) regions of the cerebral cortex played different functional roles. But no clear functional specifications arose from the early observations. Two dramatically different views emerged during the nineteenth century.

“Organology”

The earliest specific and systematic formulation on cortical localization came in the early 1800s from the German physician Franz Joseph Gall. Following a trend early in the eighteenth century in which scientists were attempting to associate body features with aspects of personality, Gall sought to determine whether there existed variation in structure and function of the brain. He developed a theory of cortical localization, which he called organology, in which each of many independent psychological faculties is mediated by a specialized organ in the brain. The central axiom of this theory was that individual differences in specific faculties were reflected in greater development of the mediating brain organ and, corre-
Introduction

Spondingly, the size of the overlying skull area. The theory was developed using a combination of observations on individual variation in specific psychological faculties and skull areas in humans and on comparisons between the abilities and skulls of animals versus humans. Gall's detailed investigations on humans included a variety of individuals that represented extreme variations in behavioral capacities. He sought out people with a special talent, such as writers, statesmen, and musical and mathematical prodigies, or with a behavioral abnormality, such as lunatics, the feebleminded, and criminals. For each he would interview the person extensively to characterize their unusual behavioral qualities and carefully examine the head for irregularities. Based on his insights about the functional aspects of their abilities and on discovery of unusual skull features, he envisioned a tight correlation between the skull and brain anatomy and a direct link to the unusual aspect of behavior. For example, among his earliest findings on humans was the observation that some people who were outstanding in memorizing verbal material had bulging eyes, suggesting to Gall an enhanced development of an organ in the frontal lobes specialized for verbal memory.

In addition, Gall collected hundreds of skulls from animals and made detailed comparisons between the anatomical features of those skulls and those of humans. The same sort of loose correlation was applied, in this case comparing the psychological abilities of animals both between different species and with humans. From a combination of all this material he devised a system of faculties, some shared by humans and animals and some exclusively human (Fig. 1–6). An example of his reasoning comes from his deductions about the faculty of "destructiveness, carnivorous instinct, or tendency to murder," localized to an area above the ears. This designation was strongly based on a combination of observations on animals and humans. That area was larger in carnivores than in grass-eating animals. And the area was overly large in a successful businessman who gave up his profession to become a butcher, in a student who was fond of torturing animals and became a surgeon, and in a pharmacist who became an executioner.

Gall's early attempts to make functional assignments of cortical areas were considerably off-base, due to two major flaws in his methodology. First, Gall's methods were based on individual cases, each of which was subject to considerable interpretation about the nature of the basis of their unusual abilities. Second, Gall simply assumed a close correlation between skull and brain anatomy, and had little interest in examining brain directly. Gall's assumptions led him in the wrong direction in almost every case. Later clinical and experimental studies consistently failed to confirm Gall's specific functional assignments. However, these later studies would demonstrate localization of cortical functions, albeit with different functional designations. Thus, quite rightfully Gall is given considerable credit for the basic insight that the cortex is composed of multiple, functionally distinct areas.

Case studies of human patients with localized cortical damage

More compelling evidence for specific functional designations within the cerebral cortex came from observations of neurological case studies of patients with selective brain damage, and from parallel physiological studies. Among the most important of the case studies on brain pathology was one made by the French physician Paul Broca in 1861. The study involved a 51-year-old man named Leborgne who had suffered from epilepsy since birth and had later lost the power to speak and developed a right side paralysis and loss of sensitivity. This patient came to Broca's attention when he was admitted to Broca's surgical ward at the hospital for an unrelated disorder, and died within a week. The autopsy revealed a circumscribed area of damage in the third convolution of the frontal lobe on the left side of the cerebral cortex. The patient's disorder was also well circumscribed. He was virtually unable to speak—indeed, he acquired the nickname "Tan" from the only sound he made—but his mouth was not paralyzed and he retained the capacity to hear and understand speech. This case provided a compelling demonstration of a highly severe and selective behavioral disorder related to a specific cortical zone. The argument for localization of higher functions was made all the more compelling with the description of a complementary case by Carl Wernicke in 1894. In this...
case the patient was severely impaired in speech comprehension, without hearing loss or a disorder of speech production, and the damage was circumscribed to a zone within the left temporal cortex.

**Experimental neurology and neurophysiology**

The evidence from neurological cases was strongly supported by concurrent findings from experimental work on animals. The earliest studies on animals, specifically aimed to test Gall's theory, were reported in 1824 by Flourens. He failed to find localization of sensory and motor functions following cortical damage in birds. These findings and other studies that could not demonstrate selective losses in mammals became the strongest evidence against localization of cortical function. However, the use of birds and other animals with relatively less differentiated cortical areas turned out to be a poor choice for an experimental model in analyses of cortical function. The case for localization was eventually made, from studies using brain stimulation in dogs, by Gustav Fritsch and Eduard Hitzig published in 1870, and with the careful work of David Ferrier using brain lesions in monkeys, presented in 1874.

Based on earlier work showing that stimulation of the head or cortex could produce twitching movements of the musculature, Fritsch and Hitzig employed minimal levels of electrical stimulation to map the cortex of dogs. They found that stimulation of a zone within the frontal cortex resulted in specific muscle movements. Moreover, they discovered that minimal stimulation of one cortical area more anterior and dorsal produced selective movements of the forepaw on the contralateral side of the body, whereas nearby regions of stimulation resulted in muscle movements in adjacent body areas. Low-level stimulation of other cortical areas did not produce movements in any part of the body, indicating that they had isolated a specialized motor area of the cortex and that area was organized as a kind of mapping of the musculature of the body.

Based on the physiological findings of Fritsch and Hitzig, Ferrier was convinced there had to be separate cortical areas that mediated specific sensory functions, such as vision, hearing, smell, and touch, and other areas that controlled movement. He suspected that previous studies had failed to find selective behavioral-anatomical correlations because the lesions were too small or the animals selected did not have sufficiently differentiated cortical areas. He prepared two monkeys, and in each showed a selective disorder associated with a specific area of cortical damage. One monkey had a severe paralysis on the right side of the body associated with a circumscribed lesion within the left frontal area. The other monkey was completely deaf following a bilateral removal of the temporal lobe. Ferrier's evidence held up to close scrutiny by his colleagues and provided the most compelling initial evidence of distinct sensory and motor functional areas in the cerebral cortex.

This combination of studies settled the debate on localization, making it clear that the cerebral cortex does not operate as a unitary organ, but rather is composed of many functionally distinct compartments. Subsequently, the localization controversy would arise again, this time about the locus of memory traces *per se*, as distinct from the more general issue of functional localization. This time, the strong localizationist view would not hold, at least with regard to the cortex. However, a new perspective, based on anatomical considerations beyond the cortex, would show that memory is subdivided according to larger pathways or compartments that involve connections between the cortex and other brain regions.

**Consolidation**

A major topic in research of the Golden Era directly associated with the phenomenon of memory involved studies on memory performance in neurological patients with memory disorders, as well as in normal human subjects. Two phenomena of amnesia following brain damage were prominent. First, patients with memory deficits could acquire new information and remember it briefly, but showed an abnormally rapid amount and rate of forgetfulness. This phenomenon in amnesia, called anterograde amnesia, was intimately tied to the diagnosis of memory impairment, in that the disorder of memory could be contrasted with intact perception and comprehension, as well as a spared ability to hold information long enough to demonstrate the latter capacities. Second, and even more impressive, was the observation of retrograde amnesia, the loss of memories acquired before the brain trauma. Both phenomena of memory loss were systematically studied first in the Golden Era.

**The neuropathology of memory**

In 1882, the French philosopher and psychologist Theodore Ribot reviewed a large number of cases of retrograde amnesia associated with brain damage and head trauma. He observed that in those cases where memory impairment is the major consequence, memories acquired remotely before the insult were relatively preserved compared to those acquired recently just before. His formulation, which came to be known as Ribot's law, was stated as a "law of regression" by which the loss of memory is inversely related to the time elapsed between the event to be remembered and the
Introduction

injury. Ribot thus concluded that memories required a certain amount of time to be organized and fixed.

Further early systematic characterizations of memory disorders, and the incumbent insights they provide about normal memory, began with the descriptions of two forms of dementia in which memory loss plays a prominent role. In 1906, Alois Alzheimer reported on an institutionalized female patient with progressive dementia. Her first symptoms involved personality changes, but soon after she exhibited a profound memory impairment. After being shown objects and recognizing them, she immediately forgot them and the circumstances in which she had learned about them. Patients with Alzheimer's disease exhibited a set of prototypical symptoms including the cardinal signs of anterograde and retrograde amnesia. Initially, the patients would show mild memory lapses. As the disease progressed, the patients would become profoundly forgetful, remembering things said for only a few minutes and then completely losing them. Consistent with the law of regression Ribot had described for retrograde amnesia following head injury, the impairment in Alzheimer's disease was more severe for memories recently acquired than for those acquired earlier in life.

Another disease with prominent loss of memory was first described by Sergei Korsakoff in 1887. His initial report involved a group of patients with an odd combination of peripheral neuromuscular symptoms (polynévrites) and memory disorder. Many of these patients were chronic alcoholics, who came to the clinic in a global confusional state that gradually resolved, leaving an outstanding selective impairment in memory as the outstanding prominent symptom. The characterization of the memory loss in Korsakoff syndrome was similar to that for the early stage of Alzheimer's disease. These patients could follow a train of conversation, but when distracted even for a brief period, they lost both the contents of the conversation and the memory that it had taken place. The patients also showed the signs of retrograde amnesia, including the temporal gradient in which remote memories were more preserved than recent ones.

Consolidation and normal human memory

In a monograph published in 1900 Georg Muller and Alfons Pilzecker reported a large number of experiments performed on normal human subjects. They had adapted Ebbinghaus's method for learning "nonsense syllables," short and pronounceable but meaningless character strings, presenting a list of them in pairs and then asking subjects to recall the second item in each pair upon subsequent probing with the first item. A major finding involved the observations of a strong tendency for subjects to spontaneously become aware of the training pairs during the retention phase, even when they tried to suppress rehearsal. They called this phenomenon "perseveration" and linked it to the additional observation that errors made during recall involved items in the same list much more often than those from separate lists. Also, both phenomena of perseveration followed a regular time gradient—they were much more prominent for a few minutes after original learning than later. Muller and Pilzecker speculated that the perseveration reflected a transient brain activity that might play an important role in establishing and strengthening the word associations. They postulated that if this were the case, then the disruption of perseveration should have a deleterious effect on recall.

To test their hypothesis Muller and Pilzecker evaluated the effects on recall performance of interpolated material given in between the initial presentations and the recall test. They tested the effects of presenting an additional list in between training and recall on an initial list, finding that indeed recall was poorer if an additional list was presented, as compared to the results with no intervening material. They called this phenomenon "retroactive interference." They varied the nature of the interpolated material by assessing the effects of presenting pictures instead of another verbal list, and found that this distraction was also effective in producing retroactive interference. Furthermore, they varied the timing of presentation of the interpolated material, and found that delaying the distraction by more than a few minutes diminished its interfering effects considerably. These studies led them to the conclusion that there is a brain activity that normally perseverates following new learning and that this activity serves to consolidate the memory.

These findings were shortly after linked to the reports of retrograde amnesia in patients with brain insult or damage. Thus, temporally graded amnesia was explained as a disruption of the perseveration process caused by a direct functional interruption of the underlying brain activity. In 1903 William Bürhan described the effect of brain trauma as disrupting a natural physical process of organization associated with a psychological process of repetition and association, processes that required time to mature.

Succeeding decades of progress

In the following chapters we explore in greater detail all of the issues raised in this chapter. The plan for the remainder of the book is to follow up on each of the four central themes, one at a time. This might seem to suggest that these issues are entirely independent, but this is very much not the case. The discovered characteristics of conditioned reflexes guided much
of the succeeding work that unsuccessfully addressed the issue of compartmentalization of memory in the cortex. Conversely, the results of succeeding studies on the nature of cognition in memory also strongly influenced the ultimately successful advances in the compartmentalization of memory functions. And succeeding studies on both the basis of cellular connections and cognitive mechanisms have led to a more sophisticated understanding of processes underlying memory consolidation. So, while I will proceed to separate these themes as a heuristic, the research that guides them, the issues themselves, and the findings on each of them are strongly interrelated.

Part I of the book updates you on our understanding of the cellular and molecular bases of memory. Chapter 2 reviews the basic anatomy of physiology of neurons, and shows how these basic principles can be put to use in explaining how memory works in relatively simple nervous systems. Chapter 3 describes parallel successes in understanding the cellular bases of a form of plasticity characteristic of mammalian brain areas, and summarizes research indicating that this form of neural plasticity may be the fundamental mechanism of learning in many more complex brain systems.

The Part II of the book builds on the discussion of the nature of cognition in memory. I will update you on how the controversy between behaviorists and cognitivists played out in the middle of the twentieth century, and then how it was resolved by discoveries in neuroscience. In particular I consider a major discovery in the neurology of memory, a case study of amnesia that ultimately showed that memory could be isolated as a cognitive function and that laid the groundwork for resolving the controversy between cognitivist and behaviorist views of memory. Then I elaborate on our understanding of a memory system that mediates "cognitive" or, as it is called today, "declarative" memory. Chapter 4 reviews the evidence from studies of amnesia in humans, and chapter 5 covers the additional evidence from animal models of amnesia. Chapter 6 summarizes complementary evidence from observations on brain activity during declarative memory in humans and animals.

In Part III, I summarize progress on the issue of compartmentalization. In Chapter 7, I begin by describing how the controversy over cortical localization became a central issue in research on memory per se, and I show how this controversy was resolved by our modern understanding of cortical modules in information processing and memory. Then I summarize the current psychological, anatomical, and physiological evidence about multiple memory systems in the brain. Chapter 8 introduces substantial direct evidence for the existence and initial localization of multiple mem-

ory systems in the brain. Chapter 9 elaborates the anatomy and workings of the full system that mediates declarative memory. Chapters 10 and 11 elaborate on two major systems, one for procedural (habit and skill) learning and the other for emotional memory.

In Part IV, I consider progress on the issue of memory consolidation. In Chapter 12, I describe how modern research has distinguished between two different kinds of consolidation, a short-term cellular fixation process and a long-lasting reorganization process. This chapter reviews the evidence for modulation of memory fixation and considers brain mechanisms that mediate memory reorganization.

Finally, Chapter 13 returns to an emphasis on a particular part of the cerebral cortex, the prefrontal area, and how this area along with other cortical areas works to orchestrate memory. Throughout the text you will see that the issues laid out a century ago are as relevant today as when they were introduced. But now we are truly beginning to resolve the anatomy and mechanisms of memory at a level of sophistication that could not have been envisioned so long ago.

**Summing up**

There are four main themes in studies on the neurobiology of memory: 
*connection, cognition, compartmentalization, and consolidation.*

Connection concerns the most basic level of analysis of memory function, the fundamental nature of the circuitry of the brain including the elements of information processing and how they communicate with one another in the service of memory. Neurons are independent elements of information processing that are connected by synapses. In the simplest circuits, neurons connect sensory inputs to motor outputs to mediate reflex arcs. However, most reflex circuits involve more complex arrangements that offer considerable coordination and control over behavior. There are also conditioned reflexes that involve the association of an arbitrary stimulus and an unconditioned stimulus, such that the conditioned stimulus comes to produce a conditioned response that is similar in form to the unconditioned or reflexive response. Conditioning is thought to be mediated by an enhancement or elaboration of the connections between neurons involved in reflex arcs.

Cognition refers to the nature of memories at the highest level of analysis, the psychological level. The central issue in the understanding of the psychological nature of memory involves the debate between behaviorism, which espouses that all learning can be reduced to conditioned responses, and cognitivism, which argues that more complex phenomena such as insight and
inference are required to explain complex learned behavior. Over most of the twentieth century evidence for both views has been accumulated.

Compartmentalization refers to the notion that memory as a whole is distributed widely in the brain, and at the same time, there are different kinds of memory that are accomplished by specific brain modules, circuits, pathways, or systems. Gall first proposed the first detailed function mapping of the cerebral cortex, which he called “organology.” However, his flawed methods led to incorrect assignments of function–structure relations. Later neurologists discovered case studies of humans with specific cortical damage and consequent specific deficits in language. Also, physiologists demonstrated that specific cortical areas in monkeys had identifiable delimited functional roles in sensory or motor processing.

Consolidation is a hypothetical phenomenon derived from the observation that memories are initially labile and later become resistant to loss, suggesting an extended process during which memories take on a permanent form. The existence of consolidation has been shown in studies on patients with damage to the brain showing a temporally graded retrograde loss of memories, that is, intact memories for material acquired remotely prior to the damage and lost memories for materials learned recently prior to the damage. The existence of consolidation can also be observed in normal humans by interposing interfering materials briefly, but not delayed, following initial learning.

READINGS


