

Cognitive Neuroscience and the Study of Memory

Review

Brenda Milner,* Larry R. Squire,†
and Eric R. Kandel‡§

*Montreal Neurologic Institute
Montreal, Quebec H3A 2B4
Canada

†Veterans Affairs Medical Center
San Diego, California 92161
and University of California
San Diego, California 92093

‡Center for Neurobiology and Behavior
College of Physicians and Surgeons
Columbia University
Howard Hughes Medical Institute
New York, New York 10032

The neurosciences have grown rapidly over the last half century. This growth has been stimulated by two important developments. First, molecular biology has transformed cellular neurobiology and has led to a new conceptual framework for signaling, a molecular framework that encompasses not only signaling in nerve cells but in all the cells of the body. Second, work on brain and cognition, which was traditionally associated with a number of different disciplines, has merged into a single discipline: cognitive neuroscience. This has provided a new framework for the study of memory, perception, action, language, and perhaps even conscious awareness.

In this review, we will consider the second development by focusing on one aspect of cognitive neuroscience: recent progress in memory research. In so doing, we also want to consider the broader question: to what degree can these two independent and disparate strands—molecular neurobiology and cognitive neuroscience—be united? Can molecular biology enlighten the study of cognitive processes, such as learning and memory, as it has other areas of biology, such as development? In turn, can cognitive neuroscience define novel phenomena that will lead to a completely new set of molecular mechanisms and insights?

The Emergence of Cognitive Neuroscience

Cognitive neuroscience originated in two disciplines: in *psychology*, in the development of rigorous methods for analyzing behavior and cognition, and in *systems neurobiology*, in the effort to understand the structure and function of neuronal circuits of the sensory and motor systems of the brain. The fusion of these two disciplines was facilitated as well by the emergence of a coherent neuroscience—an interdisciplinary approach to the nervous system that encouraged the idea that the techniques and concepts of neurobiology and systems neuroscience might be usefully applied to the analysis of cognition.

Until the beginning of the nineteenth century, the study of normal mental activity was a part of philosophy, and the chief method for understanding the mind was introspection. By the middle of the nineteenth century, introspection began to give way to experimental approaches

that eventually led to the independent discipline of experimental psychology. In its early years, experimental psychology was concerned primarily with the study of sensation, but by the turn of the century the interests of psychologists turned to behavior itself—learning, memory, attention, perception, and voluntary action.

The development of simple experimental methods for studying learning and memory—first in humans by Hermann Ebbinghaus in 1885 and a few years later in experimental animals by Ivan Pavlov and Edgar Thorndike—led to a rigorous empirical school of psychology called *behaviorism*. Behaviorists, notably James B. Watson and Burrhus F. Skinner, argued that behavior could be studied with the precision achieved in the physical sciences, but only if students of behavior abandoned speculation about what goes on in the mind (the brain) and focused instead on *observable* aspects of behavior. For behaviorists, unobservable mental processes, especially abstractions like perception, selective attention, and memory, were deemed inaccessible to scientific study. Instead, behaviorists concentrated on examining—objectively and precisely—the relationship between specific physical stimuli and observable responses in intact animals. Their early successes in rigorously studying simple forms of behavior, including learning, encouraged them to treat all processes that intervene between the stimulus (input) and behavior (output) as *irrelevant* to a scientific study of behavior. Thus, behaviorism largely ignored mental processes. As a result, the science of behavior was defined in terms of the limited techniques used to study it. This emphasis reduced the domain of experimental psychology to a restricted set of problems, and it excluded from study some of the most fascinating features of mental life.

By the 1960s, it was not difficult for the founders of *cognitive psychology*—George Miller, Ulric Neisser, Herbert Simon, and others—to convince the scientific community of the narrowness of behaviorism. These early cognitive psychologists, building on the earlier evidence from Gestalt psychology, European neurology, and work by the British psychologist Frederic Bartlett, sought to demonstrate that our knowledge of the world is based on our biological apparatus for perceiving the world, and that perception is a *constructive* process dependent not only on the information inherent in a stimulus but also on the mental processing of the perceiver. Thus, cognitive psychology was concerned not simply with specifying the input and output for a particular behavior but also with analyzing the process by which sensory information is transformed into perception and action—that is, with evaluating how a stimulus leads to a particular behavioral response. In redirecting scientific attention to mental operations, cognitive psychologists focused on *information processing*, on the flow of sensory information from sensory receptors to its eventual use in memory and action. It was implicit in the cognitive approach to behavior that each perceptual or motor act has an *internal representation* in the brain: a representation of information in patterns of neural activity.

Once cognitive psychologists acknowledged that internal representations are an essential component of

§To whom correspondence should be addressed.

behavior, they had to come to grips with the fact that most mental processes were still largely inaccessible to experimental analysis. Without direct access to the neural substrates of internal representations it was difficult, if not impossible, to understand the path from perception to action. At about this time, the work of Vernon Mountcastle on somatic sensation, David Hubel and Torsten Wiesel on vision, and Edward Evarts on the control of movement inaugurated the neuronal analysis of perception and voluntary action. Moreover, during the 1970s, Evarts and Mountcastle developed techniques for studying the activity of single cells in the brains of awake, behaving monkeys. In their hands, and in work that followed by Robert Wurtz, Apostolos Georgopoulos, William Newsome, and others, single-unit studies in monkeys led to the first correlations between cognitive processes (such as perception, attention, and decision making) and patterns of firing of individual cells in specific brain regions. This work changed the way behavior was studied both in experimental animals and in humans; the focus now was on the information processing in the brain that leads to behavior.

The need for greater anatomical knowledge led to a renaissance of neuroanatomy, evident in the development of new techniques for tracing connections between neurons by Sanford Palay at the NIH, Walle Nauta at MIT, Matthew and Jennifer LaVail at Harvard, and Max Cowan at Washington University. The search for new neuroanatomical methods and the need to bridge anatomy and function led to the application of neuroimaging techniques (positron emission tomography [PET] scanning and functional magnetic resonance imaging [MRI]) to cognitive problems. This major advance, pioneered by Marcus Raichle and Michael Posner and by Seiji Ogawa, Ken Kwong, and others, made it possible to relate changes in activity in large populations of neurons to specific cognitive acts in living humans. By comparing the results of cellular recordings in nonhuman primates and the results of neuroimaging in humans, it has become possible to study directly the neural correlates of sensory processing, motor actions, and cognitive processes.

In the 1960s and 1970s, there was also renewed interest in the traditional discipline of neuropsychology. Early students of brain and behavior like Karl Lashley and Donald Hebb used the term neuropsychology broadly to encompass studies of experimental animals as well as studies of humans. In this sense, cognitive neuroscience is the modern forum for the same topics and issues that engaged Lashley and Hebb earlier in this century. Studies of patients with brain injury or disease that affects mental function have always been a vital part of neuropsychology, and such studies formed one of the foundations of cognitive neuroscience.

As first clearly shown for language by Pierre Paul Broca in 1863, patients with lesions of specific regions of the brain exhibit quite specific cognitive deficits. Following Broca and Wernicke, the neuropsychological attempt at regional localization remained strong in Europe and in Canada but was in good part neglected in the United States, with the exception of the work of Arthur Benton, Hans-Lukas Teuber, and Norman Geschwind. As we shall see, continuing study of the behavioral consequences of brain lesions proved to be a rich source

of information about the organization and anatomy of higher functions, including memory. Lesion studies have shown that cognition is not unitary but that there are several cognitive systems, each with independent information-processing modules. For example, the visual system of primates, a prototypical cognitive system, has specialized anatomical pathways for processing information about color, form, and movement.

Finally, computational science has made a distinctive contribution to cognitive neuroscience. Computers made it possible to model the activity of large populations of neurons and to begin to test ideas about how specific components of the brain contribute to particular cognitive processes. To understand the neural organization of a complex behavior like speech, we must understand not only the properties of individual cells and pathways but also the *network properties* of functional circuits in the brain. While network properties arise from the properties of individual neurons in the network, they need not be explainable in terms of the behavior of individual cells. Computational approaches are helpful for characterizing the system as whole, for obtaining formal descriptions of what the system is capable of doing, and for determining how the interacting constituent elements account for system properties.

This review focuses on the topic of memory, but one aspect of cognitive neuroscience. We have not attempted to document fully the remarkable progress that has been achieved in our understanding of how the nervous system learns and remembers. Rather, we focus on two key components in the study of memory, as viewed through the work that the three of us have carried out with our colleagues during the past several decades. The first component is concerned with analyzing what memory is, where it is stored, and what brain systems are involved. This is the *systems problem* of memory. The second component of memory is concerned with analyzing *how* memory is stored. This is the *molecular problem* of memory.

Where Are Memories Stored?

The question of where memory is stored emerged at the beginning of the 19th century as part of the larger question—to what degree can *any* mental process be localized within the brain? The first person to address this question was Franz Joseph Gall, who made two major conceptual contributions. First, Gall attempted to abolish mind–brain dualism. He argued, based on his anatomical studies, that the brain is the organ of the mind. Second, he appreciated that the cerebral cortex is not homogenous but contains distinctive centers that control specific mental functions. Gall therefore proposed the idea of cortical *localization*. Gall asserted that the brain does not act as a unitary organ but is divided into at least 27 faculties (others were added later), each corresponding to a specific mental faculty. He thought that even the most abstract and complex of human traits, such as generosity and secretiveness, are localized to discrete areas of the brain.

Gall was not an experimentalist. He rejected the study of neurological lesions and the surgical manipulation of experimental animals and instead attempted to locate



Figure 1. Hebb and Penfield

D. O. Hebb (right) and Wilder Penfield (left) in 1958 on the occasion of Hebb delivering the 24th Annual Hughlings Jackson lecture at the Montreal Neurological Institute.

mental faculties by examining the surface of the skulls of individuals well endowed with particular functions. Perhaps not surprisingly, with this approach he misidentified the function of most parts of the cortex. This anatomically oriented approach to personality Gall called *organology*. Later, Gall's associate, Gaspard Spurzheim, adopted the better-known term *phrenology* to describe this approach.

Gall's ideas were subjected to experimental analysis by Pierre Flourens in France in the late 1820s. Flourens attempted to isolate the contributions of different parts of the nervous system to behavior by removing from the brains of experimental animals the functional centers identified by Gall. From these experiments, Flourens concluded that individual sites in the brain are not sufficient for specific behaviors such as sexual behavior and romantic love and that all regions of the brain—especially the cerebral hemispheres of the forebrain—participate in every mental function. He proposed that any part of the cerebral hemisphere is able to perform all the functions of the hemisphere. Injury to a specific area of the cerebral hemisphere should therefore affect all higher functions equally.

Despite the findings of Broca and Wernicke on the localization of language, the ensuing debate between cortical localization and equipotentiality in cognitive function dominated thinking about mental processes, including memory, well into the first half of the twentieth century. For example, in the period from 1920 to 1950, this dispute could be followed in the work of Karl Lashley, perhaps the dominant figure in American neuropsychology in the first half of this century. Lashley explored the surface of the cerebral cortex in the rat, systematically removing different cortical areas. In so doing, he failed repeatedly to identify any particular brain region that was special to or necessary for the storage of memory. Based on these experiments, Lashley formulated the law of *mass action*, according to which the extent of the memory defect was correlated

with the size of the cortical area removed, not with its specific location (Lashley, 1929). Many years later, with additional experimental work, it was possible to arrive at a different understanding of Lashley's famous conclusion.

Perhaps the first effective answer to Lashley came from Donald Hebb (Figure 1, right). In his book *The Organization of Behavior*, Hebb (1949) convinced many that it was possible to think seriously about the brain processes underlying memory. He developed concrete proposals based on biological facts, taking into consideration the neuronal circuitry that might contribute to memory storage. To explain Lashley's result that learning could not be localized to a single brain region, Hebb suggested that assemblies of cells work together to represent information and that these assemblies are distributed over large areas of cortex. Sufficient numbers of interconnected cells will survive most lesions to ensure that information can still be represented. The idea of a distributed memory store was far sighted. With the accumulation of additional evidence, it has become apparent that no single memory center exists, and many parts of the nervous system participate in the representation of any single event.

Hebb influenced many students and colleagues—in particular, Brenda Milner, who in 1957 described the remarkable patient H. M. (Scoville and Milner, 1957). H. M. had sustained a bilateral resection of the medial structures of the temporal lobe in 1953 to relieve severe epilepsy. It was immediately evident following the surgery that H. M. had a very profound impairment of recent memory in the apparent absence of other intellectual loss (Scoville, 1954). He could not remember what he had for breakfast, and he could not find his way around the hospital or recognize members of the hospital staff (except Scoville, whom he had known for many years). It seemed as though his life from the surgery onwards was not contributing to his store of knowledge. He was able to hold immediate impressions in his mind, but as

soon as his attention was diverted they were lost. In contrast, old memories from his childhood seemed to be intact.

In fact, the encounter with H. M. was not the first encounter with this kind of memory impairment. During the early 1950s, Wilder Penfield (Figure 1, left) began to carry out unilateral removals of parts of the frontal or temporal lobe as a treatment for patients with localized injury causing seizures. The temporal-lobe removals typically included the anterior temporal neocortex together with the uncus, amygdala, and anterior parahippocampal gyrus and hippocampus on the medial aspect of the hemisphere. Milner and Penfield found that these removals produced at most mild material-specific memory deficits that varied in kind with the side of the lesion. But, unexpectedly, Milner and Penfield encountered two patients with a severe, persistent, and generalized impairment of recent memory, following a removal limited to the left temporal lobe. Because both patients had undergone extensive preoperative testing, it was easy to establish that this was a selective impairment of memory, with no accompanying intellectual loss (Penfield and Milner, 1958). The impairment was manifested clinically as a profound anterograde amnesia, such that the experiences of daily life were forgotten as soon as the focus of attention shifted to a new topic. In addition, one patient showed a retrograde amnesia covering salient events of the preceding few months and the other showed a retrograde amnesia covering the 4 preceding years.

To account for this unexpected memory loss, Milner and Penfield (1955) hypothesized that in each case there must have been a pre-existing, but undetected, atrophic lesion in the hippocampal region of the opposite hemisphere, so that when the surgeon removed the anterior hippocampus and parahippocampal gyrus in the left hemisphere, he effectively deprived the patients of medial temporal-lobe function bilaterally. The reason that Milner and Penfield focused on the hippocampal region was that one patient, P. B., had had his temporal lobectomy in two stages, and it was only after removal of the medial structures of the temporal lobe that the memory loss was seen. Their hypothesis was confirmed 9 years later, when P. B. died of a pulmonary embolism and the autopsy findings revealed the presence of long-standing extensive right hippocampal atrophy, whereas the rest of the right temporal lobe, including the amygdala and the parahippocampal gyrus, showed no significant abnormality. In contrast, on the operated (left) side, the 22 mm of the hippocampus that remained appeared to be normal (Penfield and Mathieson, 1974).

Milner and Penfield reported these two cases at the 1955 meeting of the American Neurological Association in Chicago, and Scoville read their abstract. He called Penfield and said that he thought he had seen a similar memory disturbance in a patient of his (H. M.) in whom he had carried out a bilateral medial temporal-lobe resection, also in an attempt to control epileptic seizures. Penfield asked Milner if she would like to go down to Hartford, Connecticut to study the patient, and that is how the memory deficit in H. M. became more widely known.

Clinically, H. M.'s memory disorder appeared identical

to that of Penfield's two patients, except that it was more severe. Again, there had been no intellectual loss; in fact, H. M.'s IQ had risen postoperatively, from 104 to 117, presumably because he was having far fewer seizures. His capacity for sustained attention was also remarkable. Thus, Milner showed that he could retain the number 584 for at least 15 minutes by continuous rehearsal, combining and recombining the digits according to an elaborate mnemonic scheme, but the moment his attention was diverted by a new topic, the whole event was forgotten.

H. M.'s success in remembering a three-digit number for 15 minutes in the absence of distraction was at first sight consistent with Drachman's view that amnesics can hold a simple memorandum indefinitely provided that no interfering activity claims their attention (Drachman and Arbit, 1966). Yet it was already clear that for H. M. verbal rehearsal played a key role in this holding process. In contrast, certain simple nonverbal stimuli were forgotten by him within less than a minute. The evidence for this comes from delayed paired comparison and delayed matching studies.

In 1959, Konorski described a method for testing memory of single events, which was later adapted for work with human subjects by Stepien and Sierpinski (1960). This technique, called by Milner "delayed paired comparison," consists of presenting two stimuli in succession, separated by a short time interval. The subject must then indicate whether the second stimulus is the same as or different from the first. This means that subjects must retain an impression of the first stimulus in order to compare the second one with it. Task difficulty may be increased by lengthening the intratrial interval or by introducing an intratrial distraction. Prisko (1963; cited by Milner, 1972) used the Konorski method to demonstrate H. M.'s rapid forgetting of simple perceptual material. She sampled five different sets of stimuli (three visual and two auditory), each set constituting a separate task. The stimuli used were clicks, tones, shades of red, light flashes, and nonsense patterns. At least five values were assigned to each variable, to prevent as far as possible the use of verbal mediation to bridge the retention interval. All paired stimuli were easily discriminable at zero intratrial delay. These proved to be extremely easy tasks for normal subjects, who rarely made errors even with a 60-second delay and an interpolated distraction. In contrast, H. M. performed all tasks well at zero delay, but with increasing intratrial intervals his performance deteriorated sharply, so that at the 60-second delay scores were approaching the chance level and were not further impaired by distraction.

Subsequently, Sidman, Stoddard, and Mohr (1968) confirmed Prisko's findings, using a delayed matching-to-sample technique that allowed the plotting of discrimination gradients to show how far the subject's choice of a matching stimulus deviates from the sample stimulus as the intratrial interval lengthens. In the nonverbal form of their task, H. M. was required to indicate which one of eight ellipses matched the sample stimulus. With zero delay he chose correctly most of the time, showing a normal discrimination of axis-ratios, but with increasing delays his performance deteriorated until, at 32 seconds, the sample no longer exerted any control over his

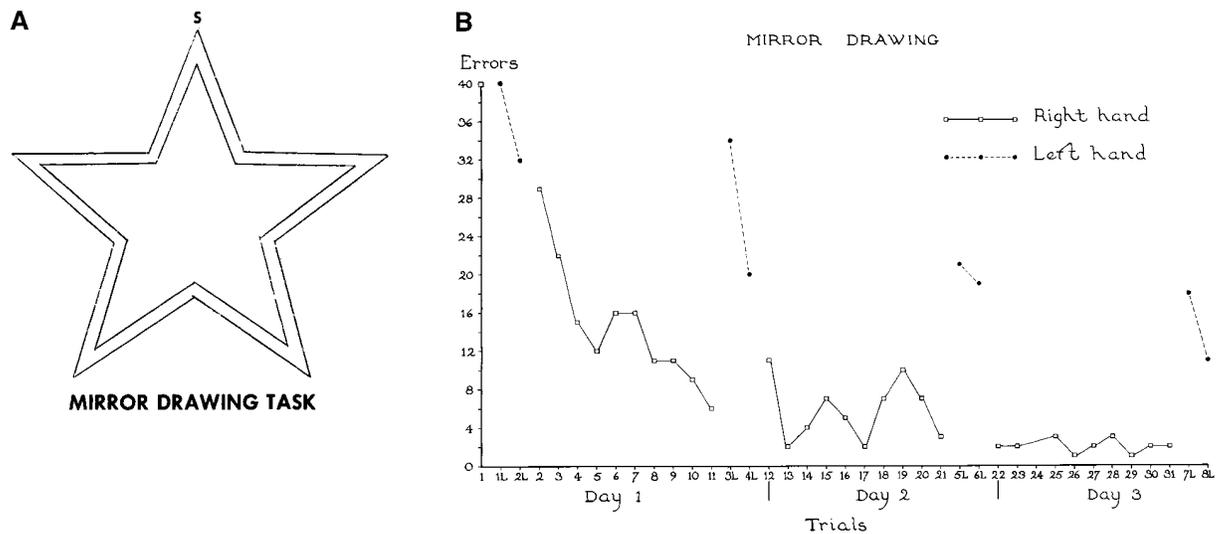


Figure 2. H. M. Showed Improvement in a Task Involving Learning Skilled Movements

In this test, he was taught to trace a line between the two outlines of a star, starting from the point S (Figure 2A), while viewing his hand and the star in a mirror. He showed steady improvement over the 3 days of testing, although he had no idea that he had ever done the task before. (The graph in Figure 2B plots the number of times, in each trial, that he strayed outside the boundaries as he drew the star.) Adapted from Milner (1962).

choice. In contrast, H. M. had no difficulty with a verbal version of the task, which required the matching of consonant trigrams. However, as with other short-term verbal memory tasks, he succeeded only by constant rehearsal; his lips could be seen moving throughout the delay period.

These and other related studies (Milner and Taylor, 1972) concur in showing that H. M. can register perceptual information normally, but that the information ceases to be available to him within about 30–40 seconds. Milner (1972) suggested that such results support the distinction between a primary memory process with a rapid decay and an overlapping secondary process (impaired in H. M.) by which the long-term storage of information is achieved.

There Are Multiple Memory Systems in the Brain

H. M.'s failure on delayed matching and delayed comparison tasks, which assess memory after a single presentation, did not rule out the possibility that he might be capable of some learning with intensive practice, or indeed that certain kinds of learning might take place at a normal rate. Accordingly, Milner and her students embarked on a variety of learning studies with H. M., including stylus maze tasks, both visual (Milner, 1965) and tactual (Corkin, 1965). With one notable exception, these studies merely served to demonstrate H. M.'s extreme difficulties with new learning, as evident also in his daily life. The exception was in the domain of motor skills, where, in 1962, Milner showed that H. M. could learn a mirror-drawing task efficiently with stable retention from day to day (Figure 2A).

If one is shown a picture of a double-margin star (Figure 2) and asked to draw a line between the two margins, one can do that very easily. However, if one has to do it while seeing one's hand and the star reflected in a mirror, then it becomes quite difficult. When one

reaches the points of the star, one tends to move the hand in the wrong direction. Eventually, with practice, we can all learn to draw the outline of a star in a mirror. It is a new sensorimotor skill, a visual-motor skill, and it is acquired across many trials. Milner was able to show that H. M. could learn that kind of task quite well. She took H. M. through 30 trials of mirror drawing spread over 3 days, and he exhibited a typical learning curve (Figure 2B). Yet at the end he had no idea he had ever done the mirror drawing task before: this was learning without any sense of familiarity. Nowadays, we are well aware that such dissociations are possible following a discrete brain lesion, but for Milner, looking at it for the first time, it was quite astonishing. Her finding contributed some of the early evidence that there is more than one memory system in the brain.

Interestingly, even before the study of patient H. M. inaugurated empirical work on the different memory systems of the brain, similar ideas had been proposed by philosophers and psychologists on the basis of intuition and introspection. For example, in 1949, Gilbert Ryle, a philosopher of mind at Oxford, proposed the existence of two types of knowledge: *knowing how*, as in knowledge of motor skills, and *knowing that*, as in the knowledge of facts and events. Some years later Jerome Bruner, one of the founders of cognitive psychology, called "knowing how" a memory *without record*. Memory without record, Bruner argued, occurs in the case of experiences that "change the nature of the organism, change his skills, or change the rules by which he operates, but are virtually inaccessible in memory as specific encounters." Here, the neural machinery that supports a behavior is presumably modified directly. He called "knowing that" a *memory with record*, a repository of information about the facts and events of everyday life.

The demonstration of intact motor skill learning in patient H. M. marked the beginning of a period of experimental work that eventually established the biological

reality of multiple memory systems. This later work made it clear that the spared memory capacities of H. M. and other amnesic patients with bilateral medial temporal-lobe lesions are not limited to motor skills. Motor skills are a subset of a large collection of learning and memory abilities, all of which are spared in amnesia and independent of the medial temporal lobe. In 1968, Warrington and Weiskrantz demonstrated what turned out to be another kind of preserved learning ability in a group of six amnesic patients, one after a right temporal lobectomy and five with alcoholic Korsakoff's psychosis. Using a version of the Gollin Figures task (Gollin, 1960), Warrington and Weiskrantz asked their patients to try to identify line drawings of common objects and animals (such as a chair or an elephant) from which most of the contour lines had been removed. This is initially quite difficult with the most fragmented drawings, but over successive presentations the contour is gradually filled in until the subject can name the item depicted. On a second presentation of the task, 1 hour later, normal subjects show considerable savings, requiring fewer contour cues to name the items. On this incomplete figures task and on an analogous fragmented words task, Warrington and Weiskrantz found marked savings in their amnesic patients, with good retention 4 weeks later, although the patients did not remember doing the tasks before. It is true that the amnesic group showed less savings than the age-matched control group, but this was only to be expected, given that the control subjects could recall most of the items and anticipated seeing them again.

Milner subsequently replicated the findings for the Gollin figures with H. M. Interestingly, H. M.'s initial performance on the first exposure to the material was above the control mean, illustrating his superior perceptual abilities. On retesting, 1 hour later, he reduced his error score by 48%, although he did not remember seeing any of the drawings before. Moreover, he showed residual savings 4 months later (Milner et al., 1968). This long-term effect of a prior visual experience, which Milner called "perceptual learning," is an instance of what is now known as priming, a form of learning distinct from motor skill and which, in this case, is probably mediated by higher visual cortical areas.

The Declarative and Nondeclarative Memory Systems

In 1980, Neal Cohen and Larry Squire showed that amnesic patients could learn the task of reading mirror-reversed print as well as normal subjects. These findings broadened further the scope of what amnesic patients could do and suggested a fundamental distinction in the way all of us process and store information about the world. The major distinction is between declarative memory and a collection of nondeclarative, nonconscious forms of memory.

Declarative memory (Figure 3) is what is ordinarily meant by the term memory. It depends on the integrity of the medial temporal lobe and affords the capacity for conscious recollections about facts and events. Declarative memory is propositional—it can be either true or false. It is involved in modelling the external world and

storing representations about facts and episodes. Nondeclarative memory is neither true nor false. It underlies changes in skilled behavior and the ability to respond appropriately to stimuli through practice, as the result of conditioning or habit learning. It also includes changes in the ability to detect or identify objects as the result of recent encounters, a phenomenon known as *priming*. In the case of nondeclarative memory, performance changes as the result of experience, which justifies the term memory, but performance changes without providing conscious access to any prior episodes (Squire et al., 1993; Schacter and Tulving, 1994). Many forms of nondeclarative memory, such as habituation, sensitization, and classical conditioning, are phylogenetically ancient and well developed in invertebrate animals that do not have a medial temporal lobe or hippocampus.

A number of nondeclarative forms of memory have been subjected to intensive study. In humans, perhaps the best studied example of nondeclarative memory is priming, first explored by Warrington and Weiskrantz (1968) and by Milner et al. (1968). Endel Tulving, Daniel Schacter, Larry Squire, and others have explored several paradigms in which subjects see lists of words, pictures of objects, or nonverbal material such as novel objects or designs (Weiskrantz, 1990; Tulving and Schacter, 1990). Subsequently, subjects are tested with both old and new items and asked to name words or objects as quickly as possible, to complete fragments to form whole items, or to make rapid decisions about items. For example, when the first few letters (MOT_) of a recently studied word (MOTEL) are presented, priming is evidenced in the tendency to complete the word fragment to form the study word instead of other possible words. Severely amnesic patients exhibit fully intact priming, despite being unable to recognize as familiar the items that had been presented previously.

Other forms of nondeclarative memory also have been studied. These include habit memory, which refers to gradually acquired dispositions or tendencies that are specific to a set of stimuli and that guide behavior. Habit learning survives hippocampal damage in humans and experimental animals but is impaired by damage to the caudate nucleus (Packard et al., 1989; Knowlton et al., 1996). Emotional learning, as in the development of phobias or in fear conditioning, is dependent on the amygdala. An enormous amount has been learned about the essential structures and connections involved in emotional learning, particularly from studies in which rats learn to fear a neutral stimulus such as a tone (fear conditioning and fear-potentiated startle) (LeDoux, 1995; Davis et al., 1997). The amygdala has also been shown to be important for emotional learning in humans (Damasio, 1995; Cahill et al., 1996). Moreover, the amygdala is essential not only for emotional learning itself; it also exerts modulatory effects on other memory systems (McGaugh et al., 1996). For example, the amygdala is responsible for the enhancement of declarative, conscious memory, which normally occurs with emotional arousal (Adolphs et al., 1997).

Perhaps the best studied example of nondeclarative memory in mammals is classical *Pavlovian* conditioning of discrete behavioral responses. A body of work initiated in the early 1980s by Richard Thompson and his

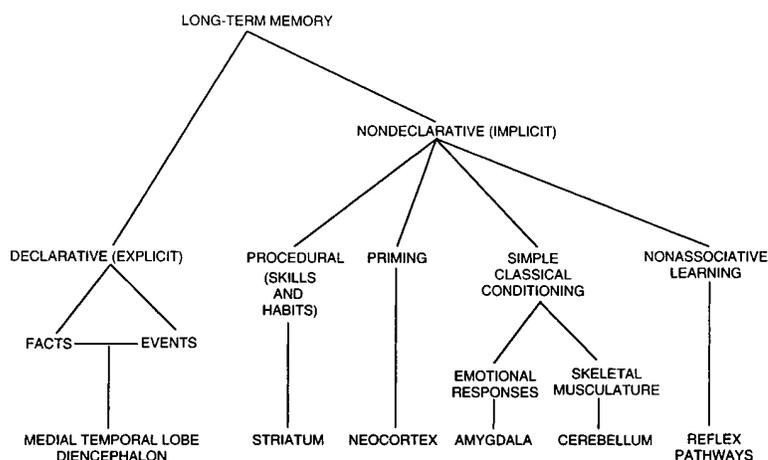


Figure 3. A Taxonomy of Mammalian Memory Systems

This taxonomy lists the brain structures and connections thought to be especially important for each kind of declarative and nondeclarative memory.

colleagues has focused on basic delay conditioning of the rabbit eyeblink response (conditioned stimulus = tone; unconditioned stimulus = airpuff; conditioned response = eyeblink). Based on anatomical findings, electrical stimulation, and reversible lesion techniques, the results provide strong evidence that the essential memory trace circuit includes the cerebellum and related brain stem circuitry and that the memory traces themselves are formed and stored in the cerebellum (Thompson and Krupa, 1994). To date, eyeblink conditioning provides the clearest information about the localization of a memory within the mammalian brain.

In humans, several kinds of nondeclarative memory have been studied, which are likely based on perceptual learning. These include adaptation-level effects, the ability to resolve random-dot stereograms, the ability to learn the regularities of "artificial grammars" by studying lawfully ordered letter strings, and the ability to acquire knowledge about categories. In category learning, one extracts and stores information about the prototype (or representative instance) of a series of items by studying many different items that, when averaged together, describe the prototype. All these forms of memory are intact in amnesic patients (Squire et al., 1993; Squire and Zola-Morgan, 1996). These kinds of memory likely involve changes within the same cortical areas responsible for perceiving and analyzing the materials that are studied.

What Parts of the Medial Temporal Lobe Are Important for Memory?

The behavioral studies reviewed above provide compelling evidence that the human declarative memory system is critically dependent upon the medial temporal region. Yet we still have much to learn about the relative importance of specific structures within the region for memory processes and the mode of interaction of these structures with other brain areas. Although Scoville and Milner (1957) drew attention to the hippocampus in the title to their paper, this was only because in their experience bilateral removals limited to the amygdala and uncus did not result in amnesia; they never claimed that the hippocampal lesions alone were responsible for H. M.'s severe memory loss.

David Amaral has recently reviewed the results of a magnetic resonance imaging study of H. M. (Corkin et al., 1997). He finds that Scoville's removal was in fact exactly as he had described it, except that the resection only extends about 5 cm posteriorly in both hemispheres, instead of the radical 8 cm originally reported. Thus, in both hemispheres the removal included the amygdala, the perirhinal and entorhinal cortex, and the anterior hippocampus. The parahippocampal cortex was largely spared. Most importantly, the temporal neocortex and the temporal stem were spared. If the roles of these various structures were to be understood, an animal model clearly was needed.

Nonhuman Primate Models of Declarative Memory

As soon as H. M. was described in 1957, efforts began to establish an animal model of his condition in the rat and monkey. If the concept of conscious recollection is central to declarative memory, how can declarative memory be studied in experimental animals? Several characteristics have been useful in extending the notion of declarative memory to mice, rats, and monkeys (Eichenbaum, 1997). These include its flexibility and the ability to use it inferentially in novel situations. It took considerable time to achieve such a model, and the first results of lesion studies in the monkey were puzzling. Animals with bilateral medial temporal-lobe resections similar to what was described in H. M. showed normal performance on visual discrimination learning tasks, even when concurrent trials on a different task were interpolated as potential "distractors" for the discrimination learning. This led many investigators to question either the human findings or the validity of cross-species comparisons. It was not until the early 1980s, with the concept of multiple memory systems and the idea that amnesia impaired only one kind of memory, that it became clearer which memory tasks were appropriate to give to experimental animals. The beginning of the solution came in 1978 when Mort Mishkin described severe deficits in monkeys with bilateral medial temporal-lobe lesions, when the monkeys were given a one-trial task of object recognition memory (delayed nonmatching to sample). This finding was consistent with the severe impairment shown by H. M. on single-trial delayed matching tasks.

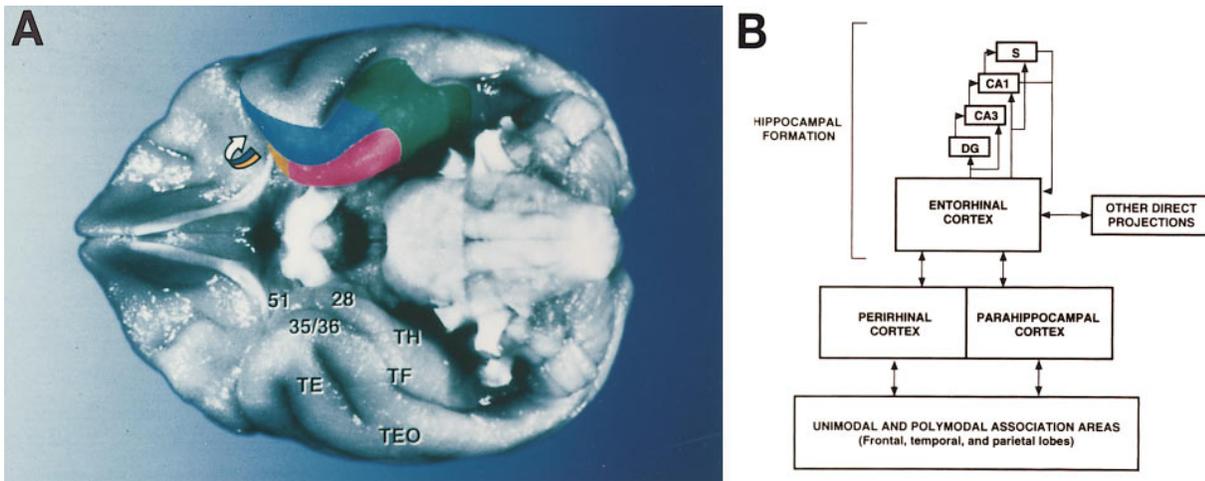


Figure 4. The Medial Temporal-Lobe Memory System in the Monkey

(A) Ventral view of a monkey brain illustrating the cortical areas underlying the hippocampus that are part of the medial temporal-lobe system. Blue, perirhinal cortex; pink, entorhinal cortex; green, parahippocampal cortex. The periamygdaloid cortex (yellow) is not thought to be a part of the system.

(B) Schematic view of the memory system. The entorhinal cortex is a major source of projections to the hippocampal region, which includes the dentate gyrus (DG), the cell fields of the hippocampus, and the subicular complex (S). Nearly two-thirds of the cortical input to entorhinal cortex originates in the adjacent perirhinal and parahippocampal cortices, which in turn receive projections from unimodal and polymodal areas in the frontal, temporal, and parietal lobes. The entorhinal cortex also receives other direct inputs from orbital frontal cortex, insular cortex, and superior temporal gyrus. All these projects are reciprocal.

The 1978 paper did not settle matters all at once but by the early 1980s, after additional work by Mishkin, Zola-Morgan, and others, an animal model of human amnesia in the monkey was established. With this model, the question of precisely which structures within the medial temporal lobe were important could be systematically explored. The identification of the anatomical components of the medial temporal-lobe memory system required about 10 years of experimental work (Squire and Zola-Morgan, 1991). The important structures are the hippocampus proper, the dentate gyrus, the subicular complex, and the entorhinal cortex (which together comprise the hippocampal formation) and the adjacent, anatomically related cortex: the perirhinal and parahippocampal cortices (Figure 4). The amygdala proved not to be a component of the declarative memory system, although it can exert a modulatory action on declarative memory.

A lesion restricted to any of the major components of this system has a significant effect on declarative memory. Indeed, two amnesic patients have been described (R. B. and G. D.) who, following an ischemic event, had bilateral lesions limited to the CA1 region of the hippocampus (Zola-Morgan et al., 1986; Rempel-Clower et al., 1996). Their deficit was qualitatively similar to H. M.'s impairment, though quantitatively it was much milder. It is now clear that the severity of H. M.'s memory impairment depends not only on his hippocampal damage but on the fact that his surgery included the hippocampal region together with the perirhinal and entorhinal cortices (Corkin et al., 1997).

A key feature of medial temporal-lobe function is that the medial temporal lobe is involved in memory for a limited period of time after learning. The initial evidence for this idea came from the observation that patient

H. M., as well as the two patients described by Penfield and Milner (1958), appeared to have intact memory for remote events that occurred years before their surgery. Subsequently, formal tests that asked about past public events also showed amnesic patients to have impaired memory for events leading up to the amnesia but intact memory for more remote events (Squire et al., 1989; Rempel-Clower et al., 1996). This loss of premonitory memory (retrograde amnesia) can cover months or even years, depending on the extent of medial temporal-lobe damage (Rempel-Clower et al., 1996).

Studies of remote memory and retrograde amnesia in amnesic patients necessarily rely on retrospective methods and imperfect tests. As a result, it is difficult to compare performance across past time periods. For these reasons, the phenomenon of retrograde amnesia has begun to be examined prospectively in experimental animals. To date, eight different studies have been carried out in which equivalent amounts of training were given at two or more times before bilateral damage to the hippocampal formation, and retention was assessed shortly after surgery (Figure 5). The work has involved mice, rats, rabbits, and monkeys and a variety of memory tasks including object discrimination learning, context-specific fear conditioning, maze learning, and trace conditioning of the eyeblink reflex. In seven of the eight studies, clear evidence was obtained for temporally graded retrograde amnesia, which covered a period ranging from a few days to about a month before surgery. In the eighth study (Bolhuis et al., 1994), memory was affected similarly at the time points tested, although performance was always at chance levels so that no difference between the two time points could have been detected.

Recent accounts of temporally graded retrograde amnesia propose that medial temporal-lobe structures

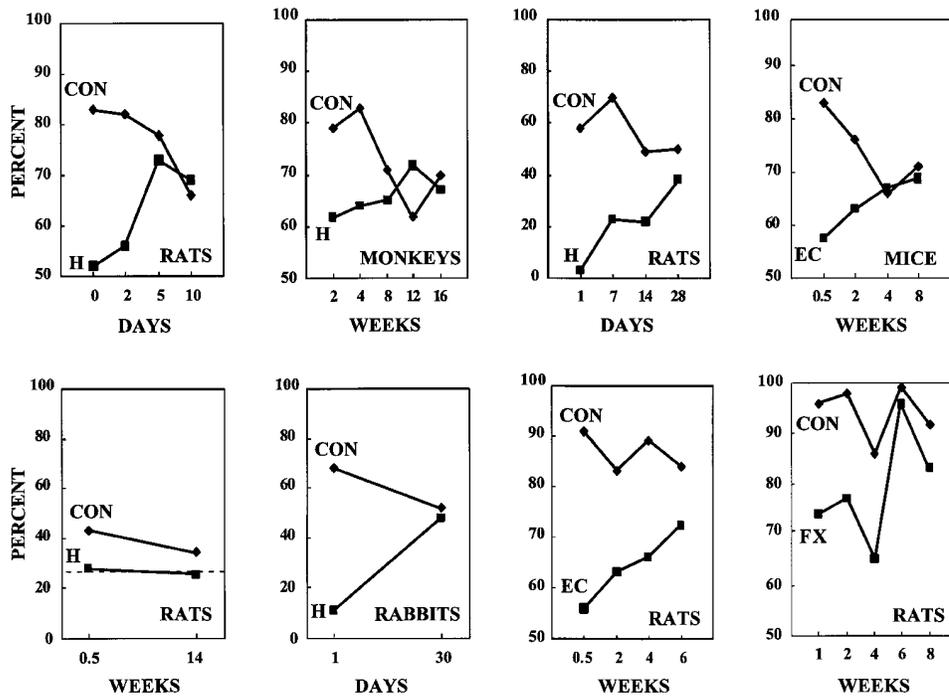


Figure 5. Summary of Findings from Eight Studies that Have Examined Retrograde Amnesia Prospectively

In these studies, an equivalent amount of training was given at each of two or more times before hippocampal formation damage, and retention was assessed shortly after surgery. In each case, the data show the performance of control (CON) and operated animals (H = hippocampus, EC = entorhinal cortex, FX = fornix) as a function of the interval between training and surgery. Control animals typically exhibited forgetting as the interval between training and surgery increased. In seven of the eight studies, operated animals exhibited temporally graded retrograde amnesia. They were impaired at retaining material they had learned recently, but they retained remotely learned material as well as control animals. In addition, the operated animals typically retained remotely learned material better than recently learned material. In the lower left panel, the dotted line denotes chance performance. From left to right, beginning on the top row, the studies are by Winocur (1990); Zola-Morgan and Squire (1990); Kim and Fanselow (1992); Cho, Beracochea, and Jaffard (1993); Bolhuis, Stewart, and Jaffard (1993); Kim, Clark, and Thompson (1995); Cho and Kesner (1996); and Wiig, Cooper, and Bear (1996).

direct a gradual process of reorganization and stabilization by changing the organization of cortical representations, for example, by gradually binding together the multiple, geographically separate cortical regions that together store memory for a whole event (Alvarez and Squire, 1994; McClelland et al., 1995). After sufficient time has passed, the hippocampal formation is not needed to support storage or retrieval of declarative memory, and long-term memory is fully dependent on the neocortex (reviewed by Squire and Alvarez, 1995).

The different components of the medial temporal lobe need not have equivalent roles in declarative memory; different structures within the medial temporal lobe are likely to carry out different subfunctions. As damage increases, fewer strategies may be available for storing memory, with the result that memory impairment becomes more severe. To study the functions of the individual regions in humans would require many patients with very specific brain lesions. Fortunately, recent anatomical and behavioral studies indicate that, even though there are differences in detail, the anatomical and functional organization of the medial temporal-lobe system is similar in humans, nonhuman primates, and simpler mammals such as rats and mice (Squire, 1992; Mayford et al., 1996). Moreover, even the mouse requires this memory system for the storage of memory about

places and objects, and this type of memory has many of the characteristics of human declarative memory, affording, for instance, the flexible use of relational information about multiple distal cues. As we shall see in the sections that follow, the possibility of studying declarative memory in mice has opened this form of memory to a molecular genetic approach.

The Molecular Biological Approach to Memory Storage

How are we to think about the cellular and molecular mechanisms of memory storage? By the end of the nineteenth century, biologists had come to appreciate that mature nerve cells have lost their capacity to divide. This fact prompted Santiago Ramón y Cajal to propose that learning does not result in the proliferation of new nerve cells but instead causes existing nerve cells to grow more branches and to strengthen their connections with other nerve cells so as to be able to communicate with them more effectively (Ramon y Cajal, 1894). This prescient idea raised three sets of questions.

First, does memory involve persistent changes in synaptic strength? If so, what are the molecular underpinnings of these synaptic changes?

Second, how do short-term synaptic changes differ from the changes that support long-term storage? Do