Back

62

Learning and Memory

Eric R. Kandel

Irving Kupfermann

Susan Iversen

BEHAVIOR IS THE RESULT OF the interaction between genes and the environment. In earlier chapters we saw how genes influence behavior. We now examine how the environment influences behavior. In humans the most important mechanisms by which the environment alters behavior are learning and memory. Learning is the process by which we acquire knowledge about the world, while memory is the process by which that knowledge is encoded, stored, and later retrieved.

Many important behaviors are learned. Indeed, we are who we are largely because of what we learn and what we remember. We learn the motor skills that allow us to master our environment, and we learn languages that enable us to communicate what we have learned, thereby transmitting cultures that can be maintained over generations. But not all learning is beneficial. Learning also produces dysfunctional behaviors, and these behaviors can, in the extreme, constitute psychological disorders. The study of learning therefore is central to understanding behavioral disorders as well as normal behavior, since what is learned can often be unlearned. When psychotherapy is successful in treating behavioral disorders, it often does so by creating an environment in which people can learn to change their patterns of behavior.

As we have emphasized throughout this book, neural science and cognitive psychology have now found a common ground, and we are beginning to benefit from the increased explanatory power that results from the convergence of two initially disparate disciplines. The rewards of the merger between neural science and cognitive psychology are particularly evident in the study of learning and memory.

In the study of learning and memory we are interested in several questions. What are the major forms of

learning? What types of information about the environment are learned most easily? Do different types of learning give rise to different memory processes? How is memory stored and retrieved?

In this chapter we review the major biological principles of learning and memory that have emerged from clinical and cognitive/psychological approaches. In the next chapter we shall examine learning and memory processes at the cellular and molecular level.

Memory Can Be Classified as Implicit or Explicit on the Basis of How Information Is Stored and Recalled

As early as 1861 Pierre Paul Broca had discovered that damage to the posterior portion of the left frontal lobe (Broca's area) produces a specific deficit in language (Chapter 1). Soon thereafter it became clear that other mental functions, such as perception and voluntary movement, can be related to the operation of discrete neural circuits in the brain. The successes of efforts to localize brain functions led to the question: Are there also discrete systems in the brain concerned with memory? If so, are all memory processes located in one region, or are they distributed throughout the brain?

In contrast to the prevalent view about the localized operation of other cognitive functions, many students of learning doubted that memory functions could be localized. In fact, until the middle of the twentieth century many psychologists doubted that memory was a discrete function, independent of perception, language, or movement. One reason for the persistent doubt is that memory storage does indeed involve many different regions of the brain. We now appreciate, however, that these regions are not equally important. There are several fundamentally different types of memory storage, and certain regions of the brain are much more important for some types of storage than for others.

The first person to obtain evidence that memory processes might be localized to specific regions of the human brain was the neurosurgeon Wilder Penfield. Penfield was a student of Charles Sherrington, the pioneering English neurophysiologist who, at the turn of the century, mapped the motor representation of anesthetized monkeys by systematically probing the cerebral cortex with electrodes and recording the activity of motor nerves. By the 1940s Penfield had begun to apply similar methods of electrical stimulation to map the motor, sensory, and language functions in the cerebral cortex of patients undergoing brain surgery for the relief of focal epilepsy. Since the brain itself does not have pain receptors, brain surgery is painless and can be carried out under local anesthesia in patients that are fully awake. Thus, patients undergoing brain surgery are able to describe what they experience in response to electrical stimuli applied to different cortical areas. On hearing about these experiments, Sherrington, who had always worked with monkeys and cats, told Penfield, “It must be great fun to put a question to the [experimental] preparation and have it answered!”

Penfield explored the cortical surface in more than a thousand patients. On rare occasions he found that electrical stimulation of the temporal lobes produced what he called an *experiential response* —a coherent recollection of an earlier experience. These studies were provocative, but they did not convince the scientific community that the temporal lobe is critical for memory because all of the patients Penfield studied had epileptic seizure foci in the temporal lobe, and the sites most effective in eliciting experiential responses were near those foci. Thus the responses might have been the result of localized seizure activity. Furthermore, the responses occurred in only 8% of all attempts at stimulating the temporal lobes. More convincing evidence that the temporal lobes are important in memory emerged in the mid 1950s from the study of patients who had undergone bilateral removal of the hippocampus and neighboring regions in the temporal lobe as treatment for epilepsy.

The first and best-studied case of the effects on memory of bilateral removal of portions of the temporal lobes was the patient called H.M., studied by Brenda Milner, a colleague of Penfield and the surgeon William Scoville. H.M., a 27-year-old man, had suffered for over 10 years from untreatable bilateral temporal lobe seizures as a consequence of brain damage sustained at age 9 when he was hit and knocked over by someone riding a bicycle. As an adult he was unable to work or lead a normal life. At surgery the hippocampal formation, the amygdala, and parts of the multimodal association area of the temporal cortex were removed bilaterally (Figure 62-1).

H.M.'s seizures were much better controlled after surgery, but the removal of the medial temporal lobes left him with a devastating memory deficit. This memory deficit (or *amnesia)* was quite specific. H.M. still had normal short-term memory, over seconds or minutes. Moreover, he had a perfectly good long-term memory for events that had occurred before the operation. He remembered his name and the job he held, and he vividly remembered childhood events, although he showed some evidence of a retrograde amnesia for information acquired in the years just before surgery. He retained a perfectly good command of language, including his normally varied vocabulary, and his IQ remained unchanged in the range of bright-normal.

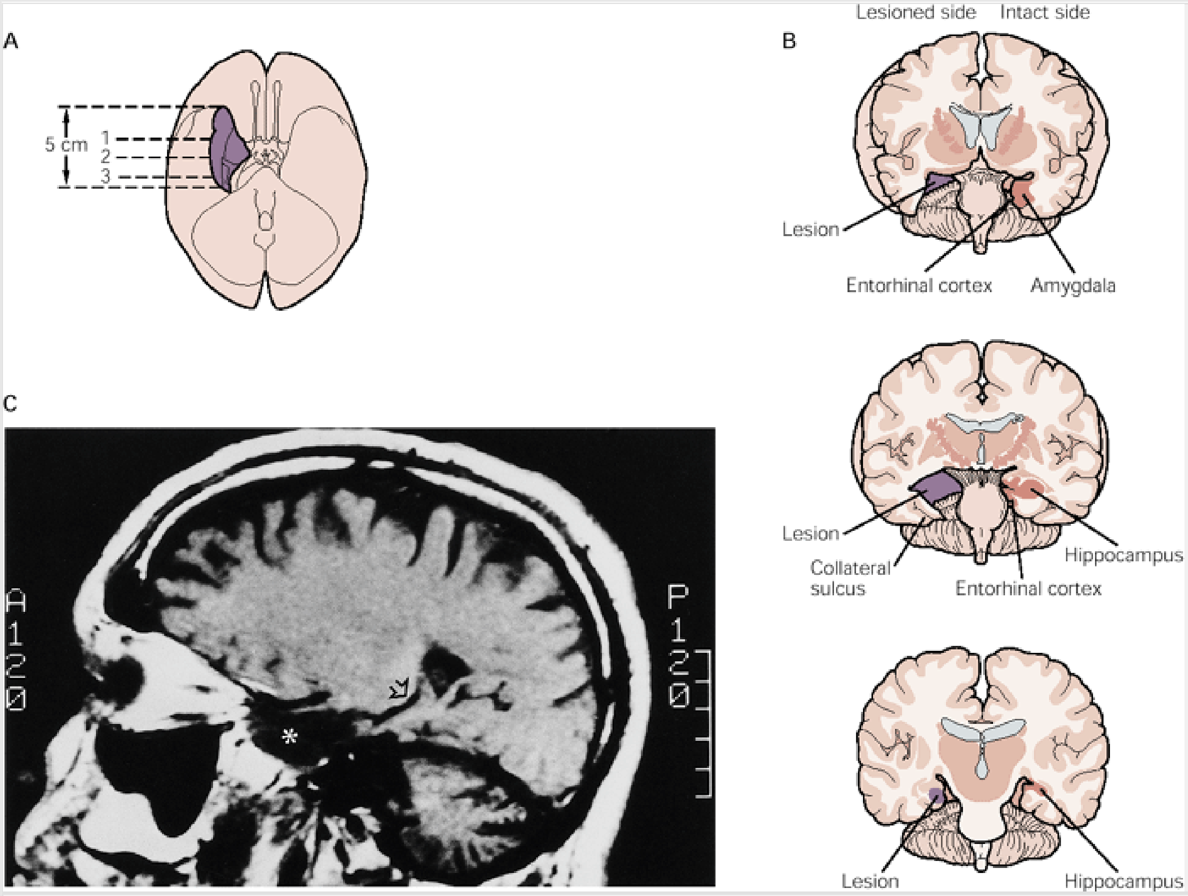


Figure 62-1 The medial temporal lobe and memory storage.

A. The longitudinal extent of the temporal lobe lesion in the patient known as H.M. in a ventral view of the brain.

B. Cross sections showing the estimated extent of surgical removal of areas of the brain in the patient H.M. Surgery was a bilateral, single-stage procedure. The right side is shown here intact to illustrate the structures that were removed. (Modified from Milner 1966.)

C. Magnetic resonance image (MRI) scan of a parasagittal section from the left side of H.M.'s brain. The calibration bar on the right side of the panel has 1 cm increments. The resected portion of the anterior temporal lobes is indicated with an **asterisk.** The remaining portion of the intraventricular portion of the hippocampal formation is indicated with an **open arrow.** Approximately 2 cm of preserved hippocampal formation is visible bilaterally. Note also the substantial cerebellar degeneration obvious as enlarged folial spaces. (From Corkin et al. 1997.)

What H.M. now lacked, and lacked dramatically, was the ability to transfer new short-term memory into long-term memory. He was unable to retain for more than a minute information about people, places, or objects. Asked to remember a number such as 8414317, H.M. could repeat it immediately for many minutes, because of his good short-term memory. But when distracted, even briefly, he forgot the number. Thus, H.M. could not recognize people he met after surgery, even when he met them again and again. For example, for several years he saw Milner on an almost monthly basis, yet each time she entered the room H.M. reacted as though he had never seen her before. H.M. had a similarly profound difficulty with spatial orientation. It took him about a year to learn his way around a new house. H.M. is not unique. All patients with extensive bilateral lesions of the limbic association areas of the medial temporal lobe, from either surgery or disease, show similar memory deficits.

The Distinction Between Explicit and Implicit Memory Was First Revealed With Lesions of the Limbic Association Areas of the Temporal Lobe

Milner originally thought that the memory deficit after bilateral medial temporal lobe lesions affects all forms of memory equally. But this proved not to be so. Even though patients with lesions of the medial temporal lobe have profound memory deficits, they are able to learn certain types of tasks and retain this learning for as long as normal subjects. The spared component of memory was first revealed when Milner discovered that H.M. could learn new motor skills at a normal rate. For example, he learned to draw the outlines of a star while looking at his hand and the star in a mirror (Figure 62-2). Like normal subjects learning this task, H.M. initially made many mistakes, but after several days of training his performance was error-free and indistinguishable from that of normal subjects.

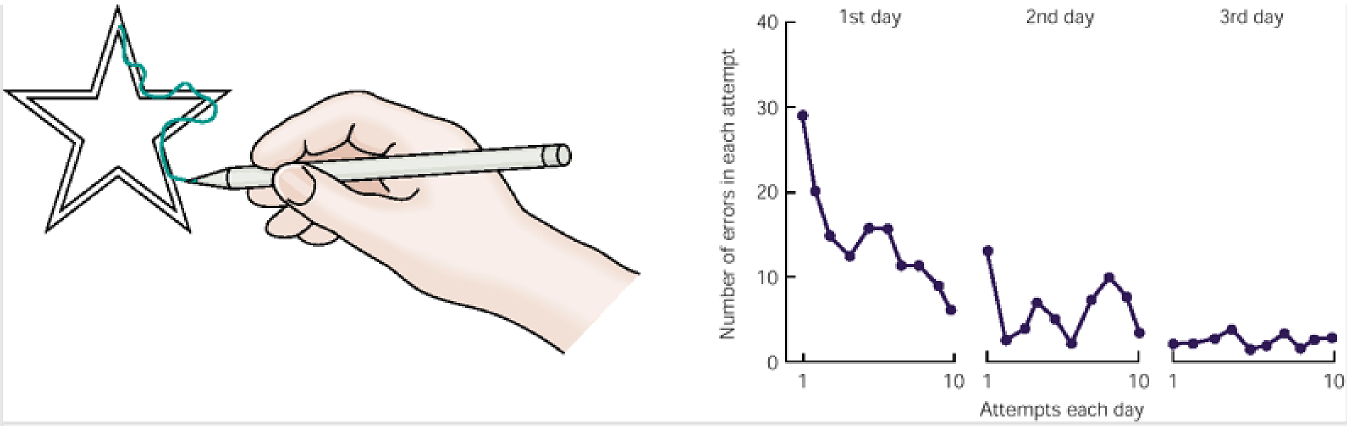


Figure 62-2 The patient H.M. showed definite improvement in any task involving learning skilled movements. He was taught to trace between two outlines of a star while viewing his hand in a mirror. He improved considerably with each fresh test, although he had no recollection that he had ever done the task before. The graph plots the number of times, in each trial, that he strayed outside the outlines as he drew the star. (From Blakemore 1977.)

Later work by Larry Squire and others has made it clear that the memory capacities of H.M. and other patients with bilateral medial temporal lobe lesions are not limited to motor skills. Rather, these patients are capable of various forms of simple reflexive learning, including habituation, sensitization, classical conditioning, and operant conditioning, which we discuss later in this chapter. Furthermore, they are able to improve their performance on certain perceptual tasks. For example, they do well with a form of memory called *priming*, in which the recall of words or objects is improved by prior exposure to the words or object. Thus, when shown the first few letters of previously studied words, a subject with amnesia correctly selects as many previously presented words as do normal subjects, even though the subject has *no* conscious memory of having seen the word before (Figure 62-3)

The memory capability that is spared in patients with bilateral lesions of the temporal lobe typically involves learned tasks that have two things in common. First, the tasks tend to be reflexive rather than reflective in nature and involve habits and motor or perceptual skills. Second, they do not require conscious awareness or complex cognitive processes, such as comparison and evaluation. The patient need only respond to a stimulus or cue, and need not try to remember anything. Thus, when given a highly complex mechanical puzzle to solve the patient may learn it as quickly and as well as a normal person, but will not consciously remember having worked on it previously. When asked why the performance of a task is much better after several days of practice than on the first day, the patient may respond, “What are you talking about? I've never done this task before.”

Although these two fundamentally different forms of memory—for skills and for knowledge—have been demonstrated in detail in amnesia patients with lesions of the temporal lobe, they are not unique amnesiacs. Cognitive psychologists had previously distinguished these two types of memory in normal subjects. They refer to information about how to perform something as *implicit memory* (also referred to as *nondeclarative memory*), a memory that is recalled unconsciously. Implicit memory is typically involved in training reflexive motor or perceptual skills. Factual knowledge of people, places, and things, and what these facts mean, is referred to as *explicit memory* (or *declarative memory*). This is recalled by a deliberate, conscious effort (Figure 62-4). Explicit memory is highly flexible and involves the association of multiple bits and pieces of information. In contrast, implicit memory is more rigid and tightly connected to the original stimulus conditions under which the learning occurred

The psychologist Endel Tulving first developed the idea that explicit memory can be further classified as *episodic* (a memory for events and personal experience) or *semantic* (a memory for facts). We use episodic memory when we recall that we saw the first flowers of spring yesterday or that we heard Beethoven's *Moonlight Sonata*

several months ago. We use semantic memory to store and recall objective knowledge, the kind of knowledge we learn in school and from books. Nevertheless, all explicit memories can be concisely expressed in declarative statements, such as “Last summer I visited my grandmother at her country house” (episodic knowledge) or “Lead is heavier than water” (semantic knowledge).

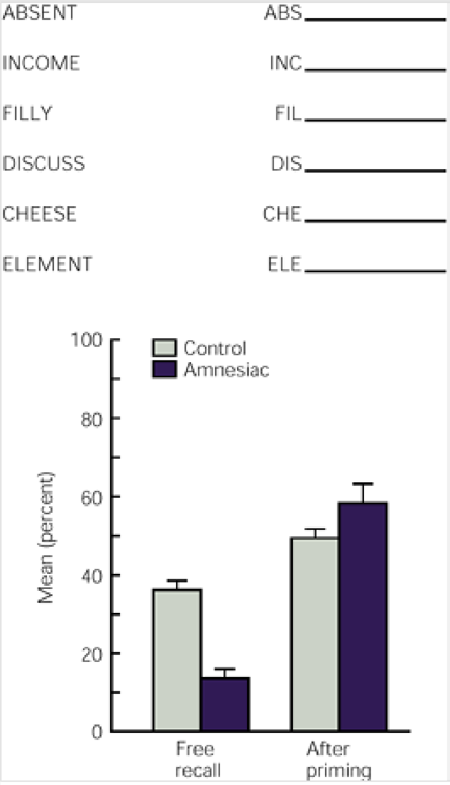


Figure 62-3 In a study of recall of words, amnesiacs and normal control subjects were tested under two conditions. First they were presented with common words and then asked to recall the words (free recall). Amnesiac patients were impaired in this condition. However, when subjects were given the first three letters of a word and instructed simply to form the first word that came to mind (completion), the amnesiacs performed as well as normal subjects. The baseline guessing rate in the word completion condition was 9%. (From Squire 1987.)

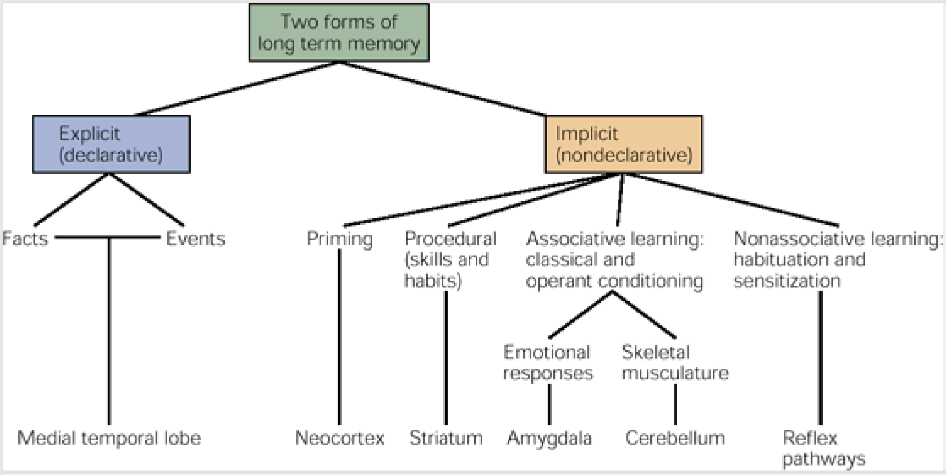

Two forms of long term memory
Explicit (declarative)
Implicit (nondeclarative)
Facts
Events
Priming
Procedural (skills and habits)
Associative learning: classical and operant conditioning
Nonassociative learning: habituation and sensitization
Emotional responses
Skeletal musculature
Medial temporal lobe
Neocortex
Striatum
Amygdala
Cerebellum
Reflex pathways

Figure 62-4 Various forms of memory can be classified as either explicit (declarative) or implicit (nondeclarative).

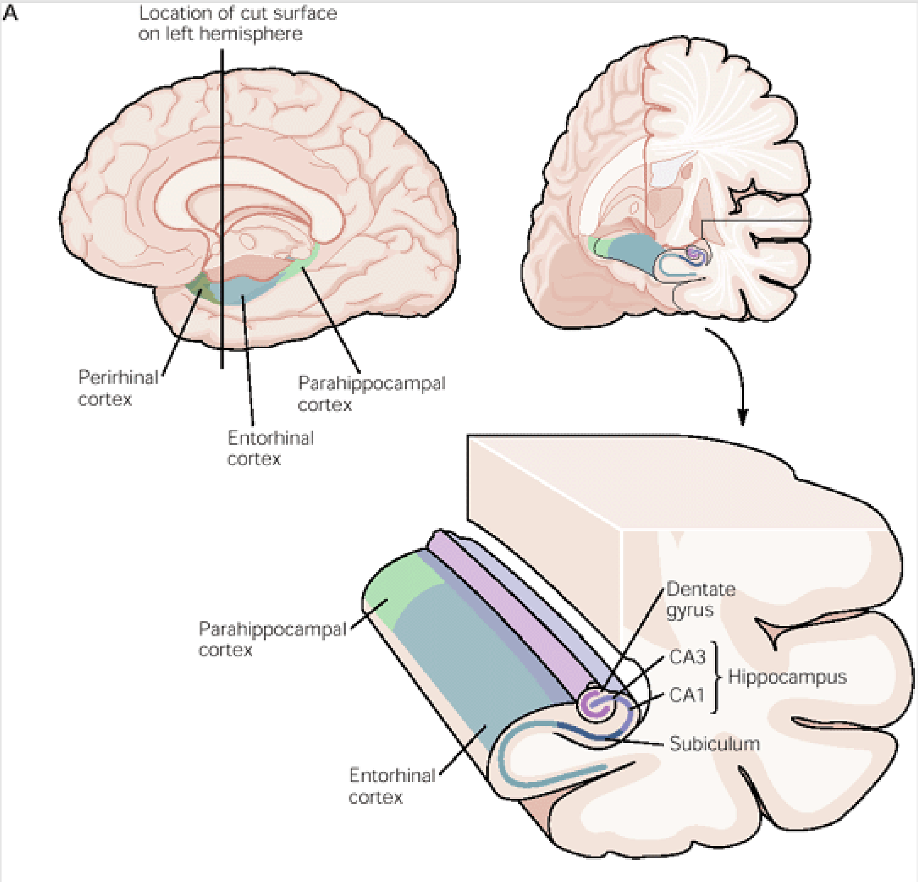
Animal Studies Help to Understand Memory

The surgical lesion of H.M.'s temporal lobe encompassed a number of regions, including the temporal pole, the ventral and medial temporal cortex, the amygdala, and the hippocampal formation (which includes the hippocampus proper, the subiculum, and the dentate gyrus) as well as the surrounding entorhinal, perirhinal, and parahippocampal cortices. Since lesions restricted to any one of these several sectors of the medial temporal lobe are rare in humans, experimental lesion studies in monkeys have helped define the contribution of the different parts of the temporal lobe to memory formation.

Mortimer Mishkin and Squire produced lesions in monkeys identical to those reported for H.M. and found defects in explicit memory for places and objects similar to those observed in H.M. Damage to the amygdala alone had no effect on explicit memory. Although the amygdala stores components of memory concerned with emotion (Chapter 50), it does not store factual information. In contrast, selective damage to the hippocampus or the polymodal association areas in the temporal cortex with which the hippocampus connects—the perirhinal and parahippocampal cortices—produces clear impairment of explicit memory.

Thus, studies with human patients and with experimental animals suggest that knowledge stored as explicit memory is first acquired through processing in one or more of the three polymodal association cortices (the prefrontal, limbic, and parieto-occipital-temporal cortices) that synthesize visual, auditory, and somatic information. From there the information is conveyed in series to the parahippocampal and perirhinal cortices,

then the entorhinal cortex, the dentate gyrus, the hippocampus, the subiculum, and finally back to the entorhinal cortex. From the entorhinal cortex the information is sent back to the parahippocampal and perirhinal cortices and finally back to the polymodal association areas of the neocortex (Figure 62-5).



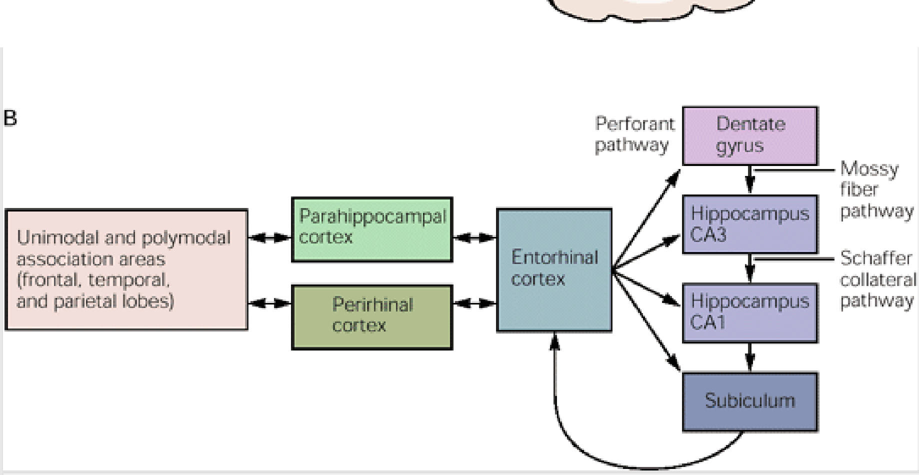


Figure 62-5 The anatomical organization of the hippocampal formation.

A. The key components of the medial temporal lobe important for memory storage can be seen in the medial (**left**) and ventral (**right**) surface of the cerebral hemisphere.

B. The input and output pathways of the hippocampal formation.

Thus, in processing information for explicit memory storage the entorhinal cortex has dual functions. First, it is the main input to the hippocampus. The entorhinal cortex projects to the dentate gyrus via the perforant pathway and by this means provides the critical input pathway through which the polymodal information from the association cortices reaches the hippocampus (Figure 62-5B). Second, the entorhinal cortex is also the major output of the hippocampus. The information coming to the hippocampus from the polymodal association cortices and that coming from the hippocampus to the association cortices converge in the entorhinal cortex. It is therefore understandable why the memory impairments associated with damage to the entorhinal cortex are particularly severe and why this damage affects not simply one but all sensory modalities. In fact, the earliest pathological changes in Alzheimer disease, the major degenerative disease that affects explicit memory storage, occurs in the entorhinal cortex.

Damage Restricted to Specific Subregions of the Hippocampus Is Sufficient to Impair Explicit Memory Storage

Given the large size of the hippocampus proper, how extensive does a bilateral lesion have to be to interfere with explicit memory storage? Clinical evidence from several patients, as well as studies in experimental animals, suggests that a lesion restricted to *any* of the major components of the system can have a significant effect on memory storage. For example Squire, David Amaral, and their collegues found that the patient R.B. had only one detectable lesion after a cardiac arrest—a destruction of the pyramidal cells in the CA1 region of the hippocampus. Nevertheless, R.B. had a defect in explicit memory that was qualitatively similar to that of H.M., although quantitatively it was much milder.

The different regions of the medial temporal lobe may, however, not have equivalent roles. Although the hippocampus is important for object recognition, for example, other areas in the medial temporal lobe may be even more important. Damage to the perirhinal, parahippocampal, and entorhinal cortices that spares the underlying hippocampus produces a greater deficit in memory storage, such as object recognition than do selective lesions of the hippocampus that spare the overlying cortex

On the other hand, the hippocampus may be relatively more important for spatial representation. In mice and rats lesions of the hippocampus interfere with memory for space and context, and single cells in the hippocampus encode specific spatial information (Chapter 63). Moreover, functional imaging of the brain of normal human subjects shows that spatial memories involve more intense hippocampal activity in the right hemisphere than do memories for words, objects, or people, while the latter involve greater activity in the hippocampus in the dominant left hemisphere. These physiological findings are consistent with the finding that lesions of the right hippocampus give rise to problems with spatial orientation, whereas lesions of the left hippocampus give rise to defects in verbal memory (Figure 62-6).

Explicit Memory Is Stored in Association Cortices

Lesions of the medial temporal lobe in patients such as H.M. and R.B. interfere only with the long-term storage of *new* memories. These patients retain a reasonably good memory of earlier events, although with severe lesions such as those of H.M. there appears to be some retrograde amnesia for the years just before the operation. How does this come about?

The fact that patients with amnesia are able to remember their childhood, the lives they have led, and the factual knowledge they acquired before damage to the hippocampus suggests that the hippocampus is only a temporary way station for long-term memory. If so, long-term storage of episodic and semantic knowledge would occur in the unimodal or multimodal association areas of the cerebral cortex that initially process the sensory information

For example, when you look at someone's face, the sensory information is processed in a series of areas of the cerebral cortex devoted to visual information, including the unimodal visual association area in the inferotemporal cortex specifically concerned with face recognition (see Box 28-1 and Chapter 28). At the same time, this visual information is also conveyed through the mesotemporal association cortex to the parahippocampal, perirhinal, and entorhinal cortices, and from there through the perforant pathway to the hippocampus. The hippocampus and the rest of the medial temporal lobe may then act, over a period of days or weeks, to facilitate storage of the information about the face initially processed by the visual association area of the inferotemporal lobe. The cells in the visual association cortex concerned with faces are interconnected with other regions that are thought to store additional knowledge about the person whose face is seen, and these connections could also be modulated by the hippocampus. Thus the hippocampus might also serve to bind together the various components of a richly processed memory of a person.

Viewed in this way the hippocampal system would mediate the initial steps of long-term storage. It would then slowly transfer information into the neocortical storage system. The relatively slow addition of information to the neocortex would permit new data to be stored in a way that does not disrupt existing information. If the association areas are the ultimate repositories for explicit memory, then damage to association cortex should destroy or impair recall of explicit knowledge that is acquired before the damage. This is in fact what happens. Patients with lesions in association areas have difficulty in recognizing faces, objects, and places in their familiar world. Indeed, lesions in different association areas give rise to specific defects in either semantic or episodic memory.

Semantic (Factual) Knowledge Is Stored in a Distributed Fashion in the Neocortex

As we have seen, semantic memory is that type of long-term memory that embraces knowledge of objects, facts, and concepts as well as words and their meaning. It includes the naming of objects, the definitions of spoken words, and verbal fluency.

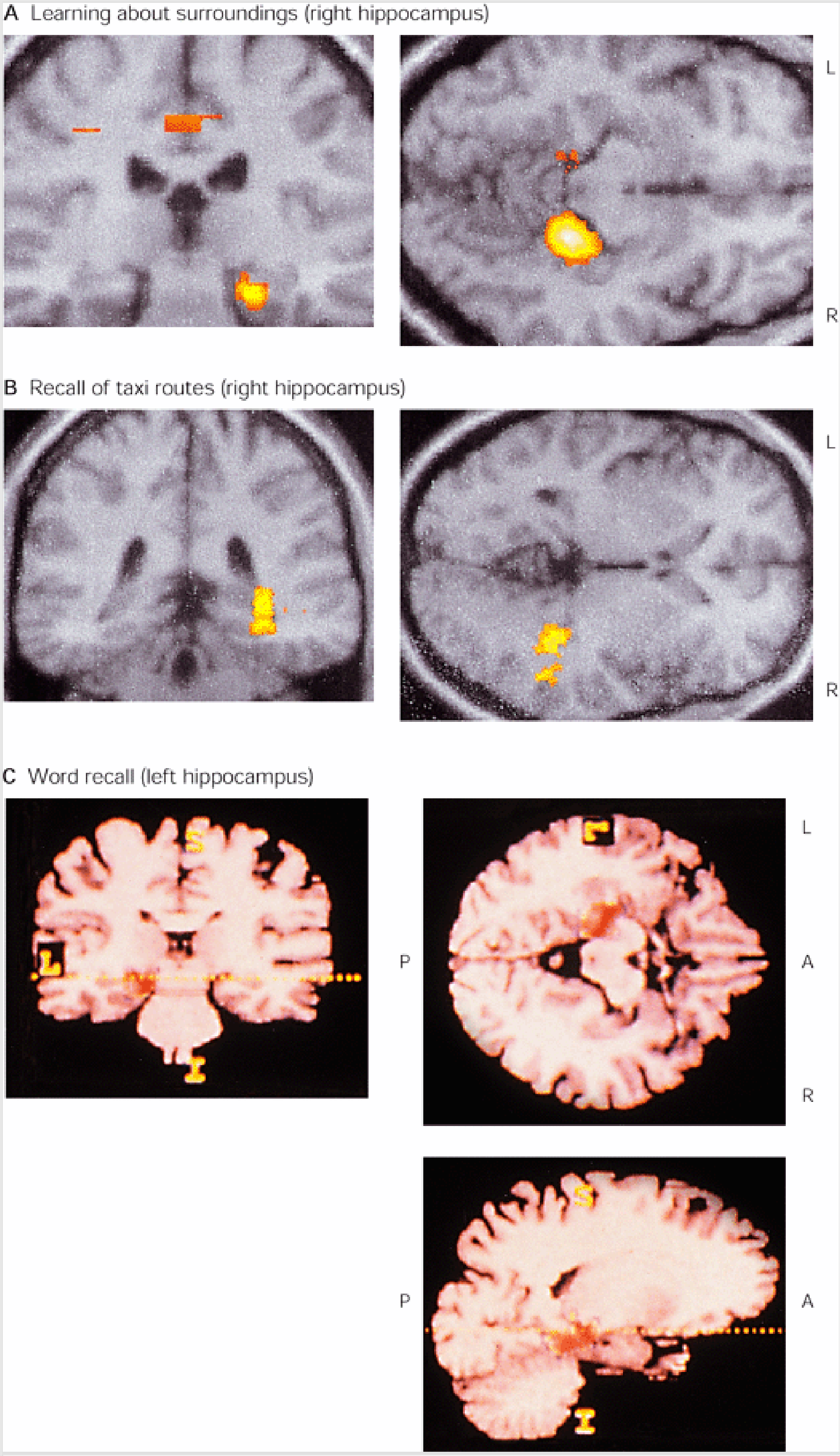

A Learning about surroundings (right hippocampus)
L
R
B Recall of taxi routes (right hippocampus)
L
R
C Word recall (left hippocampus)
S
L
I
L
P
A
R
S
P
A
I

Figure 62-6 The role of the hippocampus in memory. We spend much of our time actively moving around our environment. This requires that we have a representation in our brain of the external environment, a representation that can be used to find our way around. The right hippocampus seems to be importantly involved in this representation, whereas the left hippocampus is concerned with verbal memory.

A. The right hippocampus is activated during learning about the environment. These scans were made while subjects watched a film that depicted navigation through the streets of an Irish town. The activity during this task was compared with that in the control task where the camera was static and people and cars came by it. In the latter case there was no learning of spatial relationships and the hippocampus was not activated. Areas with significant changes in activity, indexed by local perfusion change, are indicated in **yellow** and **orange.** The scan on the left is a coronal section and the scan on the right is a transaxial section; in each panel the front of the brain is on the right and the occipital lobe on the left. (From Maguire et al. 1996.)

B. The right hippocampus also is activated during the recall by licensed taxi drivers of routes around the city of London. These people spend a long time learning the intricacies of the road network in the city and are able to describe the shortest routes between landmarks as well as the names of the various streets. The right parahippocampal and hippocampal regions are significantly activated when they do this task. The scan on the left is a coronal section and the scan on the right is a transaxial section; in each panel the front of the brain is on the right and the occipital lobe on the left. Areas with significant changes in activity, indexed by local perfusion change, are depicted in **yellow** and **orange.** (From Maguire et al. 1996.)

C. Three anatomical slices in the coronal **(left upper)**, transverse (**right upper**), and sagittal (**right lower**) planes show activation (**red**) in the left hippocampus associated with the successful retrieval of words from long lists that have to be memorized. **A** = anterior, **P** = posterior, **I** = inferior.

How is semantic knowledge built up? How is it stored in the cortex? The organization and flexibility of semantic knowledge is both remarkable and surprising. Consider a complex visual image such as a photograph of an elephant. Through experience this visual image becomes associated with other forms of knowledge about elephants, so that eventually when we close our eyes and conjure up the image of an elephant, the image is based on a rich representation of the concept of an elephant. The more associations we have made to the image of the elephant, the better we *encode* that image, and the better we can recall the features of an elephant at a future time. Furthermore, these associations fall into different categories. For example, we commonly know that an elephant is a living rather than a nonliving thing, that it is an animal rather than a plant, that it lives in a particular environment, and that it has unique physical features and behavior patterns and emits a distinctive set of sounds. Moreover, we know that elephants are used by humans to perform certain tasks and that they have a specific name. The word *elephant* is associated with all of these pieces of information, and any one bit of information can open access to all of our knowledge about elephants.

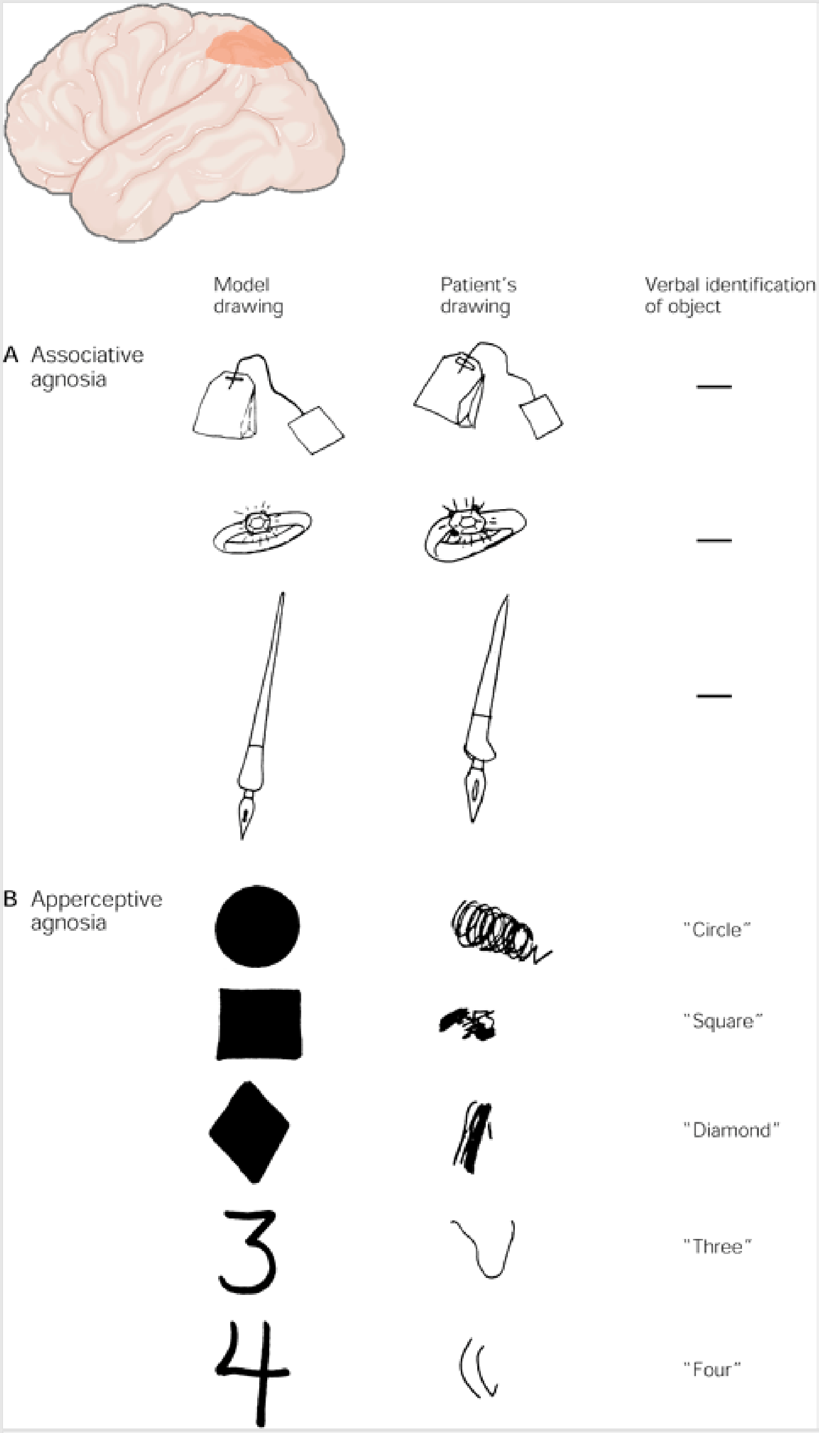


Figure 62-7 Selective lesions in the posterior parietal cortex produce selective defects in semantic knowledge. (From Farah 1994.)

A. A patient with associative agnosia is able to accurately copy drawings of a tea bag, ring, and pen but is unable to name the objects copied.

B. A patient with apperceptive agnosia is unable to reproduce even simple drawings but nevertheless can name the objects in the drawings.

As this example illustrates, we build up semantic knowledge through associations over time. The ability to recall and use knowledge—our *cognitive efficiency* —is thought to depend on how well these associations have organized the information we retain. As we first saw in Chapter 1, when we recall a concept it comes to mind in one smooth and continuous operation. However, studies of patients with damage to the association cortices have shown that different representations of an object—say, different aspects of elephants—are stored separately. These studies have made clear that our experience of knowledge as a seamless, orderly, and cross-referenced database is the product of integration of multiple representations in the brain at many distinct anatomical sites, each concerned with only one aspect of the concept that came to mind. Thus, there is no general semantic memory store; semantic knowledge is not stored in a single region. Rather, each time knowledge about anything is recalled, the recall is built up from distinct bits of information, each of which is stored in specialized *(dedicated)* memory stores. As a result, damage to a specific cortical area can lead to loss of specific information and therefore a fragmentation of knowledge.

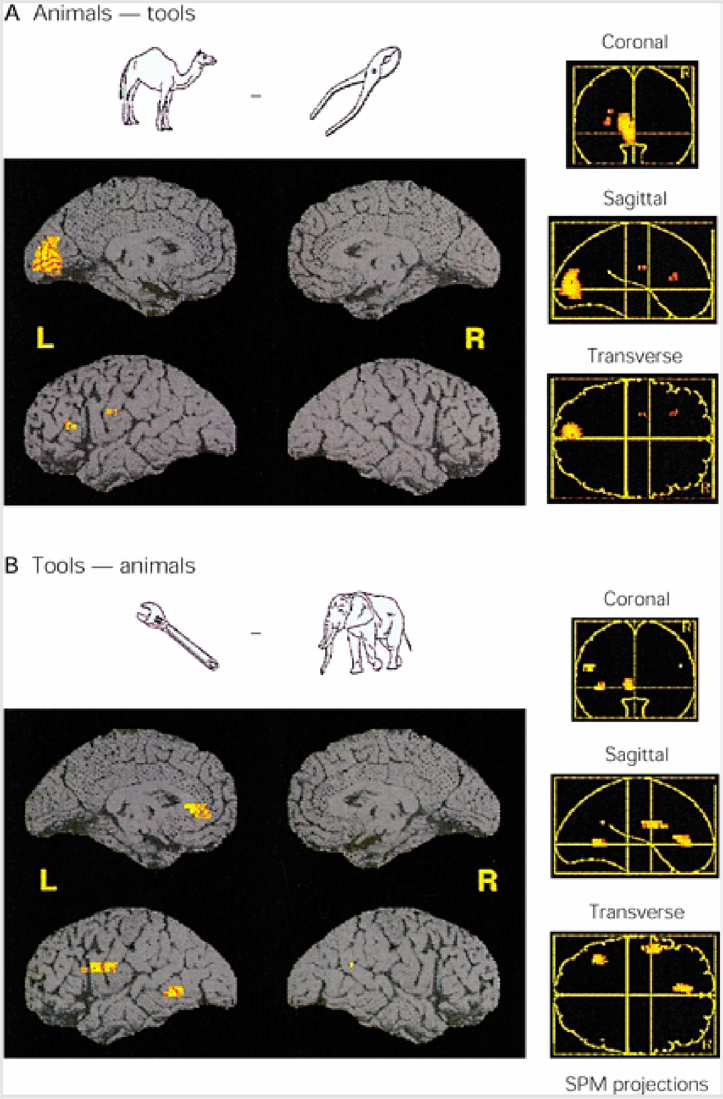


Figure 62-8 Neural correlates of category-specific knowledge are dependent on the intrinsic properties of the objects. (From Martin et al. 1996.)

A. Regions selectively activated when subjects silently named drawings of animals include the calcarine sulcus, the left and right fusiform gyri of the temporal lobe, left putamen, left thalamus, left insula and inferior frontal region (Broca's area), and right lateral and medial cerebellum. Active regions are identified by increased blood flow. In the statistical procedure regions that are active during the naming of both animals and tools cancel each other and do not show a signal. In this panel and in part B the top pair of brain images are medial views, the lower pair are lateral views. **SPM** = statistical parametric maps.

B. Regions selectively activated when subjects silently named tools are predominantly in areas of the left hemisphere associated with hand movements and the generation of action words.

For example, damage to the posterior parietal cortex can result in *associative visual agnosia;* patients cannot *name* objects but they can *identify* objects by selecting the correct drawing and can faithfully reproduce detailed drawings of the object (Figure 62-7). In contrast, damage to the occipital lobes and surrounding region can result in an *apperceptive visual agnosia;* patients are unable to draw objects but they can name them if appropriate perceptual cues are available (Figure 62-7).

While verbal and visual knowledge about objects involve different circuitry, visual knowledge involves even further specialization. For example, visual knowledge about faces and about inanimate objects is represented in different cortical areas. As we have seen in Chapter 25, lesions in the inferotemporal cortex can result in

prosopagnosia, the inability to recognize familiar faces or learn new faces, but these lesions can leave intact all other aspects of visual recognition. Conversely, positron emission tomography (PET) imaging studies show that object recognition activates the left occipitotemporal cortex and not the areas in the right hemisphere associated with face recognition. Thus, not all of our visual knowledge is represented in the same locus in the occipitotemporal cortex.

Category-specific defects in object recognition were first described by Rosaleen McCarthy and Elizabeth Warrington. They found that certain lesions interfere with the memory (knowledge) of living objects but not with memory of inanimate, manufactured objects. For example, one patient's verbal knowledge of living things was greatly impaired. When asked to define “rhinoceros” the patient responded by merely saying “animal.” But when shown a picture of a rhinoceros he responded, “enormous, weighs over a ton, lives in Africa.” The same patient's semantic knowledge of inanimate things was readily accessible through both verbal and visual cues. For example, when asked to define “wheelbarrow” he replied, “The thing we have here in the garden for carrying bits and pieces; wheelbarrows are used by people doing maintenance here on your buildings. They can put their cement and all sorts of things in it to carry it around.”

To investigate further the neural correlates of categoryspecific knowledge for animate and inanimate objects, Leslie Ungerleider and her colleagues used PET scanning to map regions of the normal brain that are associated with naming animals and regions that are involved in naming of tools. They found that naming of animals and tools both involved bilateral activation of the ventral temporal lobes and Broca's area. In addition the naming animals selectively activated the left medial temporal lobe, a region involved in the earlier stages of visual processing. In contrast, the naming tools selectively activated a left premotor area, an area also activated with hand movements, as well as an area in the left middle temporal gyrus that is activated when action words are spoken. Thus, the brain regions active during object identification are dependent in part on the intrinsic properties of the objects presented (Figure 62-8).

Episodic (Autobiographical) Knowledge About Time and Place Seems to Involve the Prefrontal Cortex

Whereas some lesions to multimodal association areas interfere with semantic knowledge, others interfere with the capacity to recall any episodic event experienced more than a few minutes previously, including dramatic personal events such as accidents and deaths in the family that occurred before the trauma. Remarkably, patients with loss of episodic memory still have the ability to recall vast stores of factual (semantic) knowledge. One patient could remember all personal facts about his friends and famous people, such as their names and their characteristics, but could not remember any specific events involving these individuals.

The areas of the neocortex that seem to be specialized for long-term storage of episodic knowledge are the association areas of the frontal lobes. These prefrontal areas work with other areas of the neocortex to allow recollection of when and where a past event occurred (Chapter 19). A particularly striking symptom in patients with frontal lobe damage is their tendency to forget how information was acquired, a deficit called *source amnesia.* Since the ability to associate a piece of information with the time and place it was acquired is at the core of how accurately we remember the individual episodes of our lives, a deficit in source information interferes dramatically with the accuracy of recall of episodic knowledge.

Explicit Knowledge Involves at Least Four Distinct Processes

We have learned three important things about episodic and semantic knowledge. First, there is not a single, all-purpose memory store. Second, any item of knowledge has multiple representations in the brain, each of which corresponds to a different meaning and can be accessed independently (by visual, verbal, or other sensory clues). Third, both semantic and episodic knowledge are the result of at least four related but distinct types of processing: encoding, consolidation, storage, and retrieval (Figure 62-9).

Encoding refers to the processes by which newly learned information is attended to and processed when first encountered. The extent and nature of this encoding are critically important for determining how well the learned material will be remembered at later times. For a memory to persist and be well remembered, the incoming information must be encoded thoroughly and deeply. This is accomplished by attending to the information and associating it meaningfully and systematically with knowledge that is already well established in memory so as to allow one to integrate the new information with what one already knows. Memory storage is stronger when one is well motivated.

Consolidation refers to those processes that alter the newly stored and still labile information so as to make it more stable for long-term storage. As we shall learn in the next chapter, consolidation involves the expression of genes and the synthesis of new proteins, giving rise to structural changes that store memory stably over time.

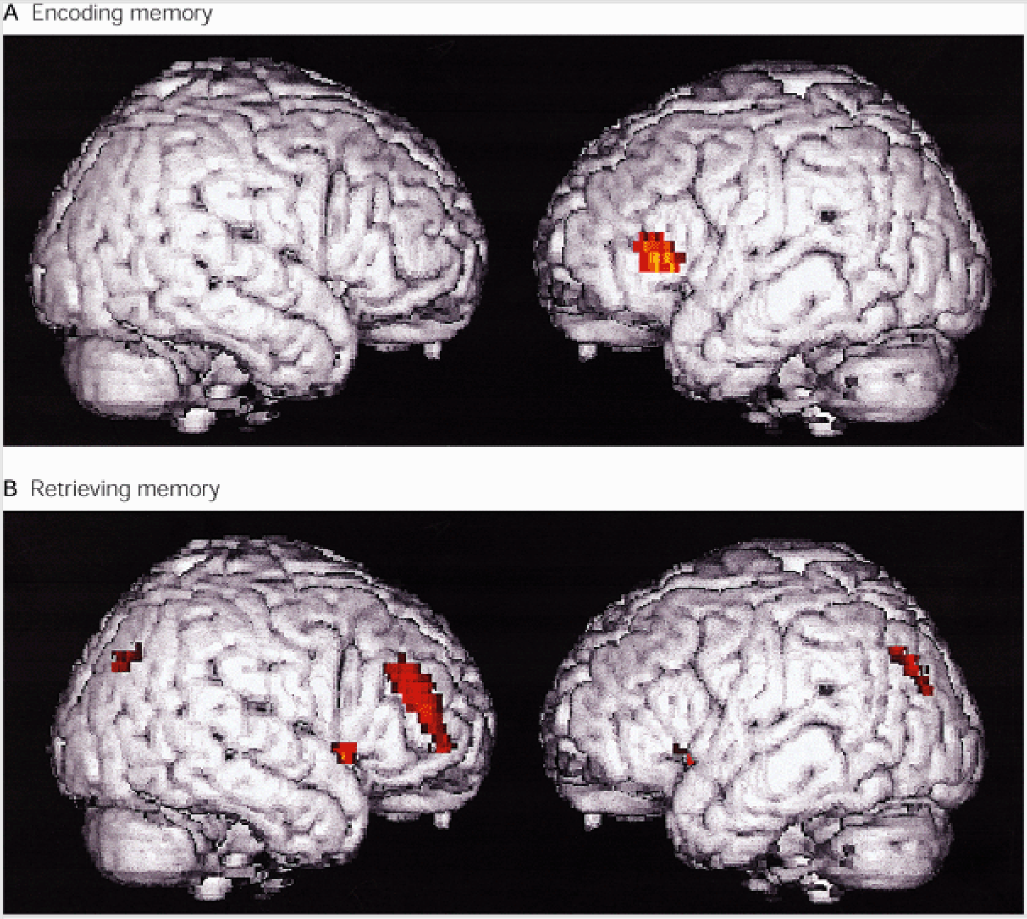


Figure 62-9 Encoding and retrieving episodic memories. Areas where brain activity is significantly increased during the performance of specific memory tasks are shown in **orange** and **red** on surface projections of the human brain (left hemisphere on the right, right hemisphere on the left).

A. Activity in the left prefrontal cortex is particularly associated with the encoding process. Subjects were scanned while attempting to memorize words paired with category labels: country—Denmark, metal—platinum, etc.

B. Activity in the right frontal cortex is associated with retrieval. Four subjects were presented with a list of category labels and examples that were not paired with the category. The subjects were then scanned when attempting to recall the examples. In addition to right frontal activation a second posterior region in the medial parietal lobe (the precuneus) is also activated.

Storage refers to the mechanism and sites by which memory is retained over time. One of the remarkable features about long-term storage is that it seems to have an almost unlimited capacity. In contrast, short-term working memory is very limited.

Finally, *retrieval* refers to those processes that permit the recall and use of the stored information. Retrieval involves bringing different kinds of information together that are stored separately in different storage sites. Retrieval of memory is much like perception; it is a constructive process and therefore subject to distortion, much as perception is subject to illusions (Box 62-1).

Retrieval of information is most effective when it occurs in the same context in which the information was acquired and in the presence of the same cues (retrieval cues) that were available to the subject during learning. Retrieval, particularly of explicit memories, is critically dependent on short-term working memory, a form of memory to which we now turn.

Working Memory Is a Short-Term Memory Required for Both the Encoding and Recall of Explicit Knowledge

How is explicit memory recalled and brought to consciousness? How do we put it to work? Both the initial encoding and the ultimate recall of explicit knowledge (and perhaps some forms of implicit knowledge as well) are thought to require recruitment of stored information into a special short-term memory store called *working memory.* As we learned in Chapter 19, working memory is thought to have three component systems.

Box 62-1 The Transformation of Explicit Memories

How accurate is explicit memory? This question was explored by the psychologist Frederic Bartlett in one series of studies in which the subjects were asked to read stories and then retell them. The recalled stories were shorter and more coherent than the original stories, reflecting reconstruction and condensation of the original. The subjects were unaware that they were editing the original stories and often felt more certain about the edited parts than about the unedited parts of the retold story. The subjects were not confabulating; they were merely interpreting the original material so that it made sense on recall.

Observations such as these lead us to believe that explicit memory, at least episodic (autobiographical) memory, is a constructive process like sensory perception. The information stored as explicit memory is the product of processing by our perceptual apparatus. As we saw in earlier chapters, sensory perception itself is not a faithful record of the external world but a constructive process in which incoming information is put together according to rules inherent in the brain's afferent pathways. It is also constructive in the sense that individuals interpret the external environment from the standpoint of a specific point in space as well as from the standpoint of a specific point in their own history. As discussed in Chapter 25, optical illusions nicely illustrate the difference between perception and the world as it is.

Moreover, once information is stored, later recall is not an exact copy of the information originally stored. Past experiences are used in the present as clues that help the brain reconstruct a past event. During recall we use a variety of cognitive strategies, including comparison, inferences, shrewd guesses, and suppositions, to generate a consistent and coherent memory.

An attentional control system (or *central executive*), thought to be located in the prefrontal cortex (Chapter 19), actively focuses perception on specific events in the environment. The attentional control system has a very limited capacity (less than a dozen items).

The attentional control system regulates the information flow to two rehearsal systems that are thought to maintain memory for temporary use: the articulatory loop for language and the visuospatial sketch pad for vision and action. The *articulatory loop* is a storage system with a rapidly decaying memory trace where memory for words and numbers can be maintained by subvocal speech. It is this system that allows one to hold in mind, through repetition, a new telephone number as one prepares to dial it. The *visuospatial sketch pad* represents both the visual properties and the spatial location of objects to be remembered. This system allows one to store the image of the face of a person one meets at a cocktail party.

The information processed in either one of these rehearsal, working memory systems has the possibility of entering long-term memory. The two rehearsal systems are thought to be located in different parts of the posterior association cortices. Thus, lesions of the extrastriate cortex impair rehearsal of visual imagery whereas lesions in the parietal cortex impair rehearsal of spatial imagery.

Implicit Memory Is Stored in Perceptual, Motor, and Emotional Circuits

Unlike explicit memory, implicit memory does not depend directly on conscious processes nor does recall require a conscious search of memory. This type of memory builds up slowly, through repetition over many trials, and is expressed primarily in performance, not in words. Examples of implicit memory include perceptual and motor skills and the learning of certain types of procedures and rules.

Different forms of implicit memory are acquired through different forms of learning and involve different brain regions. For example, memory acquired through fear conditioning, which has an emotional component, is thought to involve the amygdala. Memory acquired through operant conditioning requires the striatum and cerebellum. Memory acquired through classical conditioning, sensitization, and habituation—three simple forms of learning we shall consider later—involves charges in the sensory and motor systems involved in the learning.

Implicit memory can be studied in a variety of perceptual or reflex systems in either vertebrates or invertebrates. Indeed, simple invertebrates provide useful models for studying the neural mechanisms of implicit learning.

Implicit Memory Can Be Nonassociative or Associative

Psychologists often study implicit forms of memory by exposing animals to controlled sensory experiences. Two major procedures (or paradigms) have emerged from such studies, and these have identified two major subclasses of implicit memory: nonassociative and associative. In nonassociative learning the subject learns about the properties of a single stimulus. In associative learning the subject learns about the relationship between two stimuli or between a stimulus and a behavior.

Nonassociative learning results when an animal or a person is exposed once or repeatedly to a single type of stimulus. Two forms of nonassociative learning are common in everyday life: habituation and sensitization. *Habituation* is a decrease in response to a benign stimulus when that stimulus is presented repeatedly. For example, most people are startled when they first hear the sound of a firecracker on the Fourth of July, Independence Day in the United States, but as the celebration progresses they gradually become accustomed to the noise. *Sensitization* (or *pseudoconditioning*) is an enhanced response to a wide variety of stimuli after the presentation of an intense or noxious stimulus. For example, an animal responds more vigorously to a mild tactile stimulus after it has received a painful pinch. Moreover, a sensitizing stimulus can override the effects of habituation, a process called *dishabituation.* For example, after the startle response to a noise has been reduced by habituation, one can restore the intensity of response to the noise by delivering a strong pinch.

Sensitization and dishabituation are not dependent on the relative timing of the intense and the weak stimulus; no association between the two stimuli is needed. Not all forms of nonassociative learning are as simple as habituation or sensitization. For example, imitation learning, a key factor in the acquisition of language, has no obvious associational element.

Two forms of associative learning have also been distinguished based on the experimental procedures used to establish the learning. Classical conditioning involves learning a relationship between two stimuli, whereas operant conditioning involves learning a relationship between the organism's behavior and the consequences of that behavior.

Classical Conditioning Involves Associating Two Stimuli

Since Aristotle, Western philosophers have traditionally thought that learning is achieved through the association of ideas. This concept was systematically developed by John Locke and the British empiricist school of philosophy, important forerunners of modern psychology. Classical conditioning was introduced into the study of learning at the turn of the century by the Russian physiologist Ivan Pavlov. Pavlov recognized that learning frequently consists of becoming responsive to a stimulus that originally was ineffective. By changing the appearance, timing, or number of stimuli in a tightly controlled stimulus environment and observing the changes in selected simple reflexes, Pavlov established a procedure from which reasonable inferences could be made about the relationship between changes in behavior (learning) and the environment (stimuli). According to Pavlov, what animals and humans learn when they associate *ideas* can be examined in its most elementary form by studying the association of *stimuli*.

The essence of classical conditioning is the pairing of two stimuli. The *conditioned stimulus* (CS), such as a light, tone, or tactile stimulus, is chosen because it produces either no overt response or a weak response usually unrelated to the response that eventually will be learned. The reinforcement, or *unconditioned stimulus* (US), such as food or a shock to the leg, is chosen because it normally produces a strong, consistent, overt response (the *unconditioned response*), such as salivation or withdrawal of the leg. Unconditioned responses are innate; they are produced without learning. When a CS is followed by a US, the CS will begin to elicit a new or different response called the *conditioned response.* If the US is rewarding (food or water), the conditioning is termed *appetitive;* if the US is noxious (an electrical shock), the conditioning is termed *defensive*.

One way of interpreting conditioning is that repeated pairing of the CS and US causes the CS to become an anticipatory signal for the US. With sufficient experience an animal will respond to the CS as if it were anticipating the US. For example, if a light is followed repeatedly by the presentation of meat, eventually the sight of the light itself will make the animal salivate. Thus, classical conditioning is a means by which an animal learns to predict events in the environment.

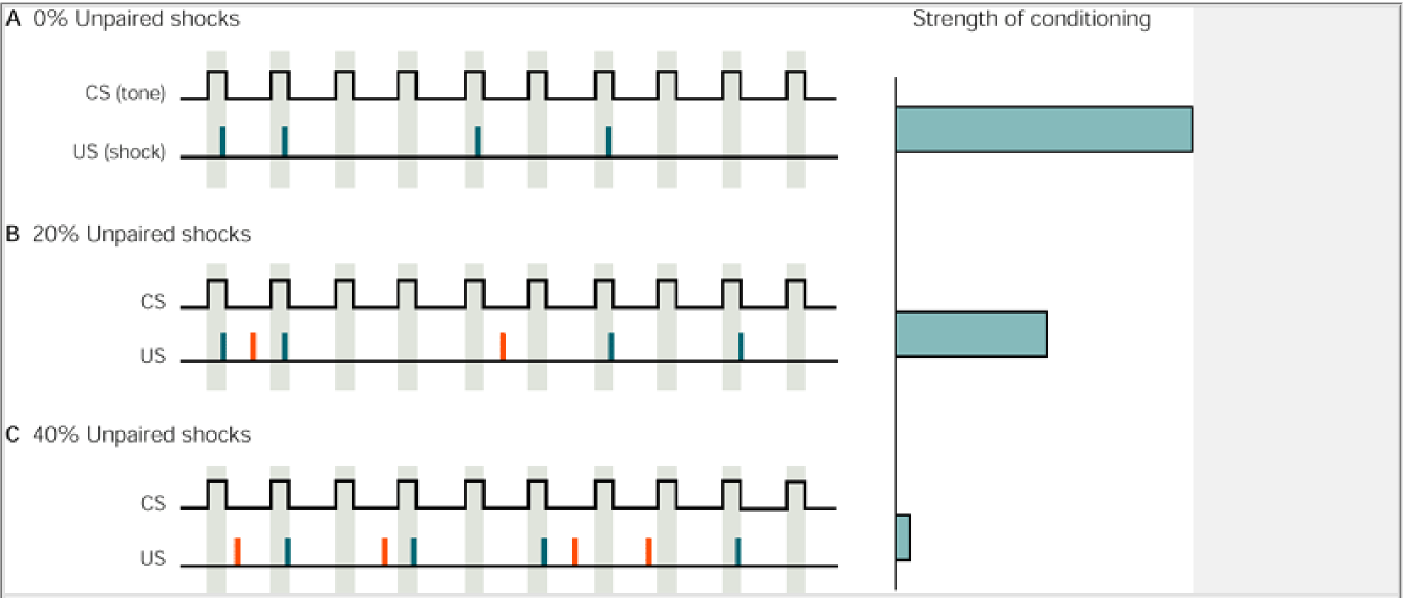


Figure 62-10 The capacity for a conditioned stimulus (CS) to produce a classically conditioned response is not a function of the number of times the CS is paired with an unconditioned stimulus (US) but rather the degree to which the CS and US are correlated. In this experiment on rats all animals were presented with a repeated tone (the CS) paired with an electric shock (the US) in 40% of the trials (**blue** vertical lines). Sometimes the shock was also presented when the tone was not present (**red** vertical lines). The percentage of these uncorrelated trials varied for different groups.

A. An experiment in which the US occurred only with the CS.

B-C. Examples in which the US was sometimes presented without the CS, either as often as the times the CS and US were paired (40%) or half as often (20%). After conditioning under the various circumstances, the degree of conditioning was evaluated by determining how effective the tone was in suppressing lever pressing to obtain food. Suppression of lever pressing is a sign of a conditioned emotional freezing response. The graph shows that in all three conditions the CS-US pairing was always 40%, but the percentage of US presentation with the absence of the CS pairing varied from 0% to 20% or 40%. When the shock occurred without the tone as often as with the tone (40%), little or no conditioning was evident. Some conditioning occurred when the shock occurred 20% of the time without the tone, and maximal conditioning occurred when the shock never occurred without the tone. (Adapted from Rescorla 1968.)

The intensity or probability of occurrence of a conditioned response decreases if the CS is repeatedly presented without the US (Figure 62-10). This process is known as *extinction.* If a light that has been paired with food is then repeatedly presented in the absence of food, it will gradually cease to evoke salivation. Extinction is an important adaptive mechanism; it would be maladaptive for an animal to continue to respond to cues in the environment that are no longer significant. The available evidence indicates that extinction is not the same as forgetting, but that instead something new is learned. Moreover, what is learned is not simply that the CS no longer precedes the US, but that the CS now signals that the US will *not* occur.

For many years psychologists thought that classical conditioning required only contiguity, that the CS precede the US by a critical minimum time interval. According to this view, each time a CS is followed by a reinforcing stimulus or US an internal connection is strengthened between the internal representation of the stimulus and the response or between one stimulus and another. The strength of the connection was thought to depend on the number of pairings of CS and US. This theory proved inadequate, however. A substantial body of empirical evidence now indicates that classical conditioning cannot be adequately explained simply by the temporal contiguity of events (Figure 62-10). Indeed, it would be maladaptive to depend solely on temporal contiguity. If animals learned to predict one type of event simply because it repeatedly occurred with another, they might often associate events in the environment that had no utility or advantage.

All animals capable of associative conditioning, from snails to humans, seem to associate events in their environment by detecting actual *contingencies* rather than simply responding to the *contiguity* of events. Why is this faculty in humans similar to that in much simpler animals? One good reason is that all animals face common problems of adaptation and survival. Learning provides a successful solution to this problem, and once a successful biological solution has evolved it continues to be selected. Classical conditioning, and perhaps all forms of associative learning, may have evolved to enable animals to distinguish events that reliably and predictably occur together from those that are only randomly associated. In other words, the brain seems to have evolved mechanisms that can detect causal relationships in the environment, as indicated by positively correlated or associated events.

What environmental conditions might have shaped or maintained such a common learning mechanism in a wide variety of species? All animals must be able to recognize prey and avoid predators; they must search out food that is edible and nutritious and avoid food that is poisonous. Either the appropriate information can be genetically programmed into the animal's nervous system (as described in Chapter 3), or it can be acquired through learning. Genetic and developmental programming may provide the basis for the behaviors of simple organisms such as bacteria, but more complex organisms such as vertebrates must be capable of flexible learning to cope efficiently with varied or novel situations. Because of the complexity of the sensory information they process, higher-order animals must establish some degree of regularity in their interaction with the world. An effective means of doing this is to be able to detect causal or predictive relationships between stimuli, or between behavior and stimuli.

Operant Conditioning Involves Associating a Specific Behavior With a Reinforcing Event

A second major paradigm of associational learning, discovered by Edgar Thorndike and systematically studied by B. F. Skinner and others, is operant conditioning (also called *trial-and-error learning*). In a typical laboratory example of operant conditioning an investigator places a hungry rat or pigeon in a test chamber in which the animal is rewarded for a specific action. For example, the chamber may have a lever protruding from one wall. Because of previous learning as well as innate response tendencies and random activity, the animal will occasionally press the lever. If the animal promptly receives a positive reinforcer (eg, food) when it presses the level, it will subsequently press the lever more often than the spontaneous rate.

The animal can be described as having learned that among its many behaviors (for example, grooming, rearing, and walking) one behavior (lever-pressing) is followed by food. With this information the animal is likely to take the appropriate action whenever it is hungry.

If we think of classical conditioning as the formation of a predictive relationship between two stimuli (the CS and the US), operant conditioning can be considered as the formation of a predictive relationship between a stimulus (eg, food) and a behavior (eg, lever pressing). Unlike classical conditioning, which tests the responsiveness of specific reflex responses to selected stimuli, operant conditioning involves behaviors that occur either spontaneously or without an identifiable stimulus. Operant behaviors are said to be *emitted* rather than elicited; when a behavior produces favorable changes in the environment (when it is rewarded or leads to the removal of noxious stimuli) the animal tends to repeat the behavior. In general, behaviors that are rewarded tend to be repeated, whereas behaviors followed by aversive, though not necessarily painful, consequences (punishment or negative reinforcement) are usually not repeated. Many experimental psychologists feel that this simple idea, called the *law of effect*, governs much voluntary behavior.

Because operant and classical conditioning involve different kinds of association—classical conditioning involves learning an association between two stimuli whereas operant conditioning involves learning the association between a behavior and a reward—one might suppose the two forms of learning are mediated by different neural mechanisms. However, the laws of operant and classical conditioning are quite similar, suggesting that the two forms of learning may use the same neural mechanisms.

For example, timing is critical in both forms of conditioning. In operant conditioning the reinforcer usually must closely follow the operant behavior. If the reinforcer is delayed too long, only weak conditioning occurs. The optimal interval between behavior and reinforcement depends on the specific task and the species. Similarly, classical conditioning is generally poor if the interval between the conditioned and unconditioned stimuli is too long or if the unconditioned stimulus precedes the conditioned stimulus. In addition, predictive relationships are equally important in both types of learning. In classical conditioning the subject learns that a certain stimulus predicts a subsequent event; in operant conditioning the animal learns to predict the consequences of a behavior.

Associative Learning Is Not Random But Is Constrained by the Biology of the Organism

For many years it was thought that associative learning could be induced simply by pairing any two arbitrarily chosen stimuli or any response and reinforcer. More recent

studies have shown that associative learning is constrained by important biological factors.

As we have seen, animals generally learn to associate stimuli that are relevant to their survival. This feature of associative learning illustrates nicely a principle we encountered in the earlier chapters on perception. The brain is not a tabula rasa; it is capable of perceiving some stimuli and not others. As a result, it can discriminate some relations between things in the environment and not others. Thus, not all reinforcers are equally effective with all stimuli or all responses. For example, animals learn to avoid certain foods (called *bait shyness*, because animals in their natural environment learn to avoid bait foods that contain poisons). If a distinctive taste stimulus (eg, vanilla) is followed by a negative reinforcement (eg, nausea produced by a poison), an animal will quickly develop a strong aversion to the taste. Unlike most other forms of conditioning, food aversion develops even when the unconditioned response (poison-induced nausea) occurs after a long delay (up to hours) after the CS (specific taste). This makes biological sense, since the ill effects of naturally occurring toxins usually follow ingestion only after some delay.

For most species, including humans, food-aversion conditioning occurs only when taste stimuli are associated with subsequent illness, such as nausea and malaise. Food aversion develops poorly, or not at all, if the taste is followed by a nociceptive, or painful, stimulus that does not produce nausea. Conversely, an animal will not develop an aversion to a distinctive visual or auditory stimulus that has been paired with nausea. Evolutionary pressures have predisposed the brains of different species to associate certain stimuli, or a certain stimulus and a behavior, much more readily than others. Genetic and experiential factors can also modify the effectiveness of a reinforcer in one species. The results obtained with a particular class of reinforcer vary enormously among species and among individuals within a species, particularly in humans

Food aversion may be a means by which humans ordinarily learn to regulate their diets to avoid the unpleasant consequences of inappropriate food. It may also be induced in special circumstances, as in the malaise associated with certain forms of cancer chemotherapy. Aversive conditioning to foods in the ordinary diet of patients might account in part for the depressed appetite of many patients who have cancer. The nausea that follows chemotherapy for cancer can produce aversion to foods that were tasted shortly before the treatment.

Certain Forms of Implicit Memory Involve the Cerebellum and Amygdala

Lesions in several regions of the brain that are important for implicit types of learning affect simple classically conditioned responses. The best-studied case is classical conditioning of the protective eyeblink reflex in rabbits, a specific form of motor learning. A conditioned eyeblink can be established by pairing an auditory stimulus with a puff of air to the eye, which causes an eyeblink. Richard Thompson and his colleagues found that the conditioned response (eyeblink in response to a tone) can be abolished by a lesion at either of two sites. Damage to the vermis of the cerebellum, even a region as small as 2 mm^2 abolishes the conditioned response, but does not affect the unconditioned response (eyeblink in response to a puff of air). Interestingly, neurons in the same area of the cerebellum show learning-dependent increases in activity that closely parallel the development of the conditioned behavior. Second, a lesion in the interpositus nucleus, a deep cerebellar nucleus, also abolishes the conditioned eyeblink. Thus, both the vermis and the deep nuclei of the cerebellum play an important role in conditioning the eyeblink, and perhaps other simple forms of classical conditioning involving skeletal muscle movement.

Maseo Ito and his colleagues have shown that the cerebellum is involved in another form of implicit memory. The vestibulo-ocular reflex keeps the visual image fixed by moving the eyes when the head moves (Chapter 41). The speed of movement of the eyes in relation to that of the head (the gain of the reflex) is not fixed but can be modified by experience. For example, when one first wears magnifying spectacles, eye movements evoked by the vestibulo-ocular reflex are not large enough to prevent the image from moving across the retina. With experience, however, the gain of the reflex gradually increases and the eye can again track the image accurately. As with eyeblink conditioning, the learned changes in the vestibulo-ocular reflex require not only the cerebellum (the flocculus) but also one of the deep cerebellar nuclei (the vestibular) in the brain stem (see Chapters 41 and 42). Finally, as we have seen in Chapter 50, lesions of the amygdala impair conditioned fear.

Some Learned Behaviors Involve Both Implicit and Explicit Forms of Memory

Classical conditioning, we have seen, is effective in associating an unconscious reflexive response with a particular stimulus and thus typically involves implicit forms of memory. However, even this simple form of learning may also involve explicit memory, so that the learned response is mediated at least in part by cognitive processes. Consider the following experiment. A subject lays her hand, palm down, on an electrified grill; a light (the CS) is turned on and at the same time she receives an electrical shock on one finger—she lifts her hand immediately (unconditioned response). After several light-shock conditioning trials she lifts her hand when the light alone is presented. The subject has been conditioned; but what exactly has been conditioned?

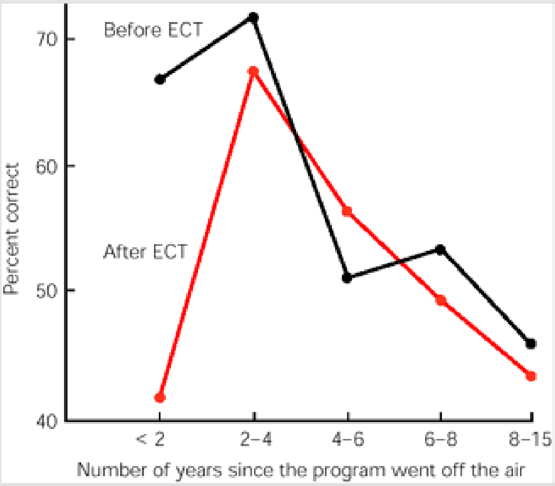


Figure 62-11 Recent memories are more susceptible than older memories to disruption by electroconvulsive treatment (ECT). The plot shows the responses of a group of patients who were tested on their ability to recall the names of television programs that were on the air during a single year between 1957 and 1972. Testing was done before and after the patients received ECT for treatment of depression. After ECT the patients showed a significant (but transitory) loss of memory for recent programs (1-2 years old) but not for older programs. (Adapted from Squire et al. 1975.)

It appears that the light is triggering a specific pattern of muscle activity (a *reflex*) that lifts the hand. However, what if the subject now places her hand on the grill, palm up, and the light is presented? If a specific reflex has been conditioned the light should produce a response that moves the hand *into* the grill. But if the subject has acquired information that the light means "grill shock," her response should be consistent with that information. In fact, the subject often will make an *adaptive response* and move her hand away from the grill. Therefore, the subject did not simply learn to apply a fixed response to a stimulus, but rather acquired information that the brain could use in shaping an appropriate response in a novel situation.

As this example makes clear, learning usually has elements of both implicit and explicit learning. For instance, learning to drive an automobile involves conscious execution of specific sequences of motor acts necessary to control the car; with experience, however, driving becomes an automatic and nonconscious motor activity. Similarly, with repeated exposure to a fact *(semantic learning*), recall of the fact with appropriate clues can eventually become virtually instantaneous—we no longer consciously and deliberately search our memory for it.

Both Explicit and Implicit Memory Are Stored in Stages

It has long been known that a person who has been knocked unconscious selectively loses memory for events that occurred before the blow (retrograde amnesia). This phenomenon has been documented thoroughly in animal studies using such traumatic agents as electroconvulsive shock, physical trauma to the brain, and drugs that depress neuronal activity or inhibit protein synthesis in the brain. Brain trauma in humans can produce particularly profound amnesia for events that occur within a few hours or, at most, days before the trauma. In such cases older memories remain relatively undisturbed. The extent of retrograde amnesia, however, varies among patients, from several seconds to several years, depending on the nature and strength of the learning and the nature and severity of the disrupting event.

Studies of memory retention and disruption of memory have supported a commonly used model of memory storage by stages. Input to the brain is processed into short-term working memory before it is transformed through one or more stages into a more permanent long-term store. A search-and-retrieval system surveys the memory store and makes information available for specific tasks.

According to this model, memory can be impaired at several points. For example, there can be a loss of the contents of a memory store; as we have seen (Chapter 58), in Alzheimer's disease there actually is a loss of nerve cells in the entorhinal cortex. Alternatively, the search-and-retrieval mechanism may be disrupted by head trauma. This latter conclusion is supported by the observation that trauma sometimes only temporarily disrupts memory, since considerable memory for past events gradually returns. If stored memory were completely lost, it obviously could not be recovered.

Studies of memory loss in patients undergoing electroconvulsive therapy (ECT) for depression have confirmed and extended the findings from animal experiments. Patients were examined using a memory test that could reliably quantify the degree of memory for relatively recent events (1-2 years old), old events (3-9 years old), and very old events (9-16 years old). The patients were asked to identify, by voluntary recall, the names of television programs broadcast during a single year between 1957 and 1972. The patients were tested before ECT and then again afterward (with a different set of television programs). Both before and after ECT recall of the programs was more correct for more recent years. After ECT, however, the patients showed a significant but transitory loss of memory for more recent programs, while their recall of older programs remained essentially the same as it was before ECT (Figure 62-11).

One interpretation of these findings is that until memories have been converted to a long-term form, retrieval (recall) of recent memories is easily disrupted.

Once converted to a long-term form, however, the memories are relatively stable. With time, however, both the long-term memory and the capacity to retrieve it gradually diminish, even in the absence of external trauma. Because of this susceptibility to disruption, the total set of retrievable memories changes continually with time.

Several experiments studying the effects of drugs on learning support the idea that memory is time-dependent and subject to modification when it is first formed. For example, subconvulsant doses of excitant drugs, such as strychnine, can improve the retention of learning of animals even when the drug is administered after the training trials. If the drug is given to the animal soon after training, retention of learning on the following day is greater. The drug has no effect, however, when given after a long delay (several hours) after training. In contrast, inhibitors of protein synthesis selectively block the formation of long-term memory but not short-term memory when given during the training procedure.

An Overall View

The neurobiological study of memory has yielded three generalizations: memory has stages, long-term memory is represented in multiple regions throughout the nervous system, and explicit and implicit memories involve different neuronal circuits.

Different types of memory processes involve different regions and combinations of regions in the brain. Explicit memory underlies the learning of facts and experiences—knowledge that is flexible can be recalled by conscious effort and can be reported verbally. Implicit memory processes include forms of perceptual and motor memory—knowledge that is stimulus-bound, is expressed in the performance of tasks without conscious effort, and is not easily expressed verbally. Implicit memory flows automatically in the doing of things, while explicit memory must be retrieved deliberately.

Long-term storage of explicit memory requires the temporal lobe system. Implicit memory involves the cerebellum and amygdala and the specific sensory and motor systems recruited for the task being learned. Moreover, the memory processes for many types of learning involve several brain structures. For example, learned changes of the vestibulo-ocular reflex appear to involve at least two different sites in the brain, and explicit learning involves neocortical structures as well as the hippocampal formation. Furthermore, there are reasons to believe that information is represented at multiple sites even within one brain structure.

This parallel processing may explain in part why a limited lesion often does not eliminate a specific memory, even a simple implicit memory. Another important factor that may account for the failure of small lesions to adversely affect a specific memory may reside in the very nature of learning. As we shall see in the next chapter, memory involves both functional and structural changes at synapses in the circuits participating in a learning task. Although such changes are likely to occur only in particular types of neurons, the complex nature of many tasks makes it likely that these neurons are widely distributed within the pathways that mediate the response. Therefore some components of the stored information (ie, some of the synaptic changes) could remain undisturbed by a small lesion. Furthermore, the brain can take even the limited store of remaining information and construct a good representation of the original, just as the brain normally constructs conscious memory.

Selected Readings

Corkin S, Amaral DG, González RG, Johnson KA, Hyman BT. 1997. H.M.'s medial temporal lobe lesion: findings from magnetic resonance imaging. J Neurosci 17:3964–3979.

Kamin LJ. 1969. Predictability, surprise, attention, and conditioning. In: BA Campbell and RM Church (eds). *Punishment and Aversive Behavior*, pp. 279-296. New York: Appleton-Century-Crofts.

Maguire EA, Frackowiak RS, Frith CD. 1996. Learning to find your way: a role for the human hippocampal formation. Proc R Soc London B 263:1745–1750.

McClelland JL, McNaughton BL, O'Reilly RC. 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psych Rev 3:419–457.

Milner B. 1966. Amnesia following operation on the temporal lobes. In: CWM Whitty and OL Zangwill (eds). *Amnesia*, pp. 109-133. London: Butterworths.

Milner B, Squire LR, Kandel ER. 1998. Cognitive neuroscience and the study of memory. Neuron 20:445–468.

Muller R. 1996. A quarter of a century of place cells. Neuron 17:813–822.

Schwartz B, Robbins SJ. 1994. *Psychology of Learning and Behavior.* 4th ed. New York: Norton.

Schacter D. 1996. *Searching For Memory. The Brain, the Mind and the Past.* New York: Harper Collins/Basic Books.

Squire LR, Kandel ER. 1999. *Memory: From Mind to Molecules.* New York: Freeman.

Squire LR, Zola-Morgan S. 1991. The medial temporal lobe memory system. Science 253:1380–1386.

Steinmetz JE, Lavond DG, Ivkovich D, Logan CG, Thompson RF. 1992. Disruption of classical eyelid conditioning after cerebellar lesions: damage to a memory trace system or a simple performance deficit? J Neurosci 12:4403–4426.

References

Bartlett FC. 1932. *Remembering: a Study in Experimental and Social Psychology.* Cambridge, England: The University Press.

Blakemore C. 1977. *Mechanics of the Mind.* Cambridge, MA: Cambridge Univ. Press.

Domjan M, Burkhard B. 1986. *The Principles of Learning and Behavior*, 2nd ed. Monterey, CA: Brooks/Cole.

Drachman DA, Arbit J. 1966. Memory and the hippocampal complex II. Is memory a multiple process? Arch Neurol 15:52–61.

du Lac S, Raymond JL, Sejnowski TJ, Lisberger SG. 1995. Learning and memory in the vestibulo-ocular reflex. Annu Rev Neurosci 18:409–441.

Farah M. 1990. *Visual Agnosia.* Cambridge, MA: MIT Press.

Frackowiak RS. 1994. Functional mapping of verbal memory and language. Trends Neurosci 17:109–115.

Hebb DO. 1966. *A Textbook of Psychology.* Philadelphia: Saunders.

Martin A, Wiggs CL, Ungerleider LG, Haxby JV. 1996. Neural correlates of category-specific knowledge. Nature 379:649–652.

McCarthy, RA, Warrington EK. 1990. *Cognitive Neuropsychology: A clinical Introduction.* San Diego: Academic Press.

McClelland JL, McNaughton BL, O'Reilly RC. 1995. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. Psychol Rev 102:419–457.

McGaugh JL. 1990. Significance and remembrance: the role of neuromodulatory systems. Psychol Sci 1:15–25.

Pavlov IP. 1927. *Conditioned Reflexes: Investigation of the Physiological Activity of the Cerebral Cortex.* Anrep GV, trans. London: Oxford University Press.

Penfield W. 1958. Functional localization in temporal and deep sylvian areas. Res Publ Assoc Res Ment Dis 36:210–226.

Rescorla RA. 1968. Probability of shock in the presence and absence of CS in fear conditioning. J Comp Physiol Psychol 66:1–5.

Rescorla RA. 1988. Behavioral studies of Pavlovian conditioning. Annu Rev Neurosci 11:329–352.

Skinner BF. 1938. *The Behavior of Organisms: An Experimental Analysis.* New York: Appleton-Century-Crofts.

Squire LR. 1987. *Memory and Brain.* New York: Oxford University Press.

Squire LR, Slater PC, Chace PM. 1975. Retrograde amnesia: temporal gradient in very long term memory following electroconvulsive therapy. Science 187:77–79.

Thorndike EL. 1911. *Animal Intelligence: Experimental Studies.* New York: Macmillan.

Tulving E, Schacter DL. 1990. Priming and human memory systems. Science 247:301–306.

Warrington EK, Weiskrantz L. 1982. Amnesia: a disconnection syndrome? Neuropsychologia 20:233–248.

NEURONAL SIGNALLING OF FEAR MEMORY

Stephen Maren and Gregory J. Quirk ‡*

Abstract | The learning and remembering of fearful events depends on the integrity of the amygdala, but how are fear memories represented in the activity of amygdala neurons? Here, we review recent electrophysiological studies indicating that neurons in the lateral amygdala encode aversive memories during the acquisition and extinction of Pavlovian fear conditioning. Studies that combine unit recording with brain lesions and pharmacological inactivation provide evidence that the lateral amygdala is a crucial locus of fear memory. Extinction of fear memory reduces associative plasticity in the lateral amygdala and involves the hippocampus and prefrontal cortex. Understanding the signalling of aversive memory by amygdala neurons opens new avenues for research into the neural systems that support fear behaviour.

If there is one central tenet of the neurobiology of learning and memory, it is that plasticity in the CNS is essential for the representation of new information. Experience-dependent plasticity in the brain might take many forms, ranging from the synthesis and insertion of synaptic proteins to whole-brain synchronization of neuronal activity. An important challenge is to understand how these various forms of experience-dependent plasticity are reflected in the activity of neuronal populations that support behaviour. Donald Hebb referred to these populations as cell assemblies, and this concept has had important heuristic value in empirical studies of the neurobiology of memory[1]. With the advent of modern electrophysiological recording techniques, Hebb's concept of the cell assembly is now amenable to experimental study in awake, freely behaving animals. Using parallel recording techniques, multiple extracellular electrodes can be used to 'listen' to the action-potential dialogue between several neurons at once[2,3] (BOX 1).

In this article, we review recent single-unit recording studies that have provided considerable insight into the neuronal mechanisms of learning and memory, focusing particularly on Pavlovian fear conditioning. In this form of learning, a neutral stimulus, such as an acoustic tone (the conditional stimulus, or CS) is paired with a noxious unconditional stimulus (US), such as a footshock. After only a few conditioning trials, the CS comes to evoke a learned fear response (conditional response, or CR). Pavlovian fear conditioning is particularly amenable to electrophysiological analysis because it is acquired rapidly and yields long-lasting memories. Moreover, the behavioural principles and neural circuits that underlie this form of learning are well characterized, allowing an unprecedented analysis of the relationship between neuronal activity and learned behaviour.

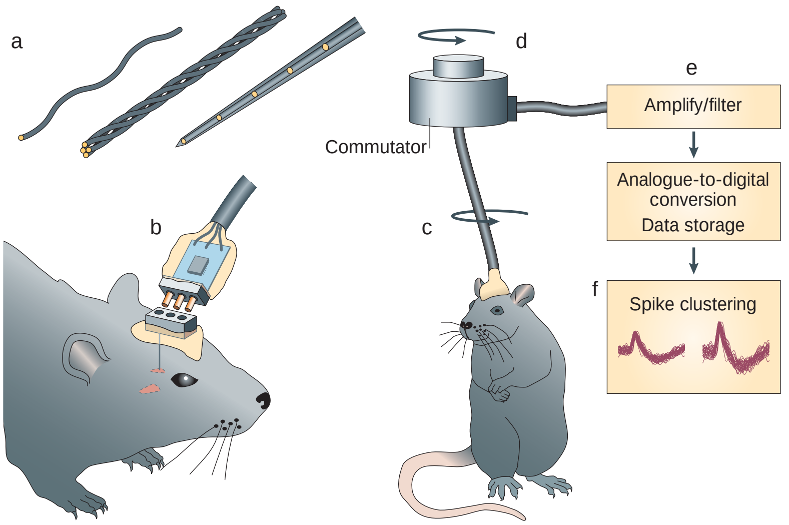
Neuronal correlates of aversive memory

The search for the neurophysiological mechanisms of aversive memory began in the early 1960s with the observation that an auditory stimulus that was paired with an electric shock modified auditory-evoked field potentials in cats and rats[4,5]. Because cortical field potentials are generated by large populations of neurons, changes in early components of the field potentials (reflecting processing in ascending auditory tracts) were variable and poorly localized. Other investigators observed changes in late components of cortical potentials that were attributed to a general state of 'fear'[6], but these changes were not associative (that is, they did not reflect a specific CS–US association) because they occurred in response to both the CS and a novel stimulus. Therefore, it became clear that field-potential recordings would not be sufficient to identify loci of fear memory.

**Department of Psychology and Neuroscience Program, University of Michigan, Ann Arbor, Michigan 48109, USA. ‡Department of Physiology, Ponce School of Medicine, Ponce 00372, Puerto Rico. Correspondence to: S.M. or G.J.Q. e-mails: maren@umich.edu; gjquirk@yahoo.com*

doi:10.1038/nrn1535

Box 1 | Single-unit recording methods



Parallel advances in computing hardware (for example, data storage capacity and processor speed), software (for example, neuronal data acquisition and spike sorting) and electrode technology have coalesced to yield powerful multichannel single-unit recording systems for behaving animals. In a typical system, recording electrodes consist of bundles of single wires, multi-wire stereotrodes or TETRODES**, or thin-film silicon arrays (a). Electrode assemblies are either chronically implanted in brain tissue or affixed to moveable microdrives, some of which have been engineered to independently drive up to 16 tetrodes (64 channels) (b). Voltages recorded on each electrode are typically passed through integrated circuits in source-follower configurations that are mounted near the animal's head (a headstage) to convert neuronal signals into low-impedance signals that are less sensitive to cable and line noise (c). Signals are then fed from the headstage through a commutator to allow free movement of the animal and cable assembly (d). Neuronal signals are amplified, band-pass filtered and digitized (e). Once digitized, spike waveforms on each electrode channel are sorted into single units using sophisticated clustering algorithms (f). The isolation of single units using such methodology varies widely and depends on several parameters. Most importantly, multichannel electrodes, such as tetrodes, seem to yield the most reliable single-unit isolation. Several commercial packages are available to acquire neuronal signals from high-density recording systems, although most electrophysiologists use a combination of home-made technology and commercial products.**

Subsequent single-unit recording studies in cats and monkeys showed conditioning-induced changes in evoked spike activity in several brain areas, including the midbrain, thalamus and cortex[7–9]. These changes seemed to be associative because they were not observed during pseudo-conditioning, in which the CS and US were unpaired. In addition, sensitizing effects of the shock were ruled out with discriminative models, in which responses to a CS that was paired with the US (CS^{+}) were compared with responses to a CS that was never paired with the US (CS^{-})[10,11]. However, from these studies it was not possible to determine whether structures that showed increased neuronal responsiveness after conditioning were primary sites of plasticity or were downstream from other plastic sites.

To address this issue, Olds and colleagues[12] assessed the latency of conditioned single-unit responses in various brain areas in an appetitive auditory conditioning task. They reasoned that structures showing the earliest increases in auditory responses (in terms of milliseconds after CS onset) were probably primary sites of plasticity, whereas those showing longer-latency changes were probably downstream sites that were involved in the expression of learned responses. Short-latency plastic responses (within 40 ms of tone onset) were observed in the posterior thalamus, medial geniculate nucleus and auditory cortex, indicating that these areas might be primary sites of plasticity. Although this approach was criticized for not taking into account descending modulation from the cortex[13], subsequent work by Disterhoft and colleagues showed that thalamic neurons were able to learn in fewer trials than cortical neurons[14,15], confirming that thalamic plasticity preceded cortical plasticity, in terms of both latency and trials.

Therefore, plasticity in subcortical structures could occur independently of the cortex, and indeed, learning-related plasticity might not even require the forebrain under some circumstances. In the most systematic neurobiological analysis of Pavlovian learning so far, Thompson and colleagues found that although hippocampal neurons show considerable plasticity during eyeblink conditioning, hippocampal plasticity is not essential for this form of learning. In fact, neuronal plasticity in the cerebellum is crucial for the acquisition and expression of eyeblink conditioning[16,17].

Fear-related plasticity in the lateral amygdala

Notably absent from these early studies of conditioning was any mention of the amygdala. The thalamus and cortex were thought to be the sites that most probably encode emotional associations (but see REF. 18), and the amygdala was suspected to have a role in modulating memory storage in these areas[19]. However, an influential study by Kapp and co-workers showed that lesions of the central nucleus of the amygdala prevented heart-rate conditioning in rabbits[20], consistent with central nucleus modulation of fear-expression centres in the midbrain and hypothalamus[21,22]. Subsequent single-unit recording studies of the central nucleus revealed associative plasticity[23,24], indicating that the amygdala might be a site of plasticity in fear conditioning.

Converging on a similar conclusion, LeDoux and co-workers discovered direct projections from the auditory thalamus to the amygdala in rats, and determined this projection to be vital for auditory fear conditioning[25–27]. Specifically, the lateral nucleus of the amygdala (LA) receives direct projections from the medial subdivision of the medial geniculate nucleus and the adjacent thalamic posterior intralaminar nucleus (MGm/PIN), and it relays this information by way of the basal amygdaloid nuclei to the central nucleus[28–31] (FIG. 1). Small lesions of the LA or the MGm/PIN prevent fear conditioning, whereas large lesions of the auditory cortex or striatum do not[32,33], indicating that thalamo–amygdala inputs are sufficient for conditioned fear responses. This finding galvanized interest in the LA as a potential site of plasticity in fear conditioning, and set the stage for the next 15 years of work on the role of the amygdala in this form of learning. Indeed, considerable research now indicates that the amygdala is necessary for both the acquisition and expression of Pavlovian fear memories[34], but not for all forms of aversive memory[35,36].

TETRODE
An extracellular electrode that comprises four juxtaposed recording channels, which can be used to disambiguate the signals emitted by individual point sources. Because each neuron occupies a unique position in space, its spikes are 'seen' slightly differently by each electrode, providing a unique signature. This technique allows the identification of many more neurons than there are sampling electrodes.

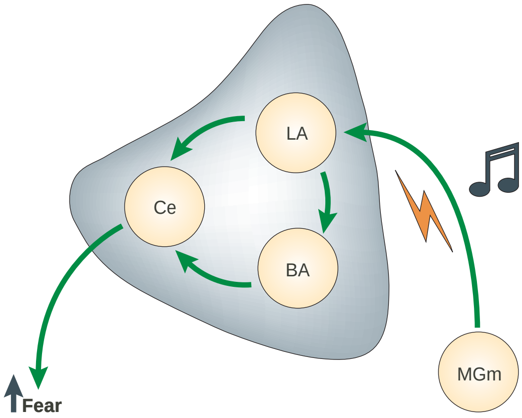


Figure 1 | **Neural circuits that are necessary for auditory fear conditioning.** Tone and shock inputs from the medial subdivision of the medial geniculate nucleus (MGm) converge in the lateral amygdala (LA), resulting in potentiation of auditory responses of LA neurons. The LA projects to the central nucleus of the amygdala (Ce), both directly and indirectly by way of the basal amygdala (BA). Descending outputs of the Ce to brainstem and hypothalamic structures trigger fear responses.

An important question is whether neurons in the LA show associative plasticity during fear conditioning. Although previous work implied that this was the case[37,38], nobody had recorded from the dorsal subdivision of the LA (LAd), which is the primary target of MGm/PIN inputs and a site of CS and US convergence. Because the LAd projects to ventral parts of the LA, which in turn project to basolateral and central nuclei, plasticity downstream from the LAd could be passively fed forward from the LAd. To address this issue, Quirk and colleagues recorded LAd neurons in behaving rats, and observed robust increases in tone responses during fear conditioning compared with a sensitization control phase[39] (FIG. 2; BOX 1). Most of the conditioned increases in spike firing occurred within 15 ms of tone onset, corresponding to the latency of thalamic (12 ms) rather than cortical (>20 ms) activation of LA neurons[40]. Maren subsequently confirmed this extremely short-latency plasticity in LAd, and showed that it persisted at these latencies through extensive overtraining[41]. Parallel work has revealed that LA neurons show synaptic LONG-TERM POTENTIATION (LTP)[42–44], and that fear conditioning is associated with LTP-like changes in thalamo–amygdala synaptic transmission[45–47]. Together with evidence of converging auditory and somatosensory inputs onto LA neurons from the thalamus[48,49], this indicated that the LAd might be a site of long-term memory in fear conditioning (BOX 2).

Although these findings are consistent with a primary locus of conditioning-related plasticity in the LAd, it is necessary to show that LAd plasticity is not passively fed forward from either the auditory thalamus or the auditory cortex. Indeed, short-latency plastic responses in fear conditioning have been observed in both the MGm/PIN[50] and the auditory cortex[51]. To determine the contribution of the cortical pathway, Quirk and colleagues compared conditioned unit responses of LAd neurons with those in

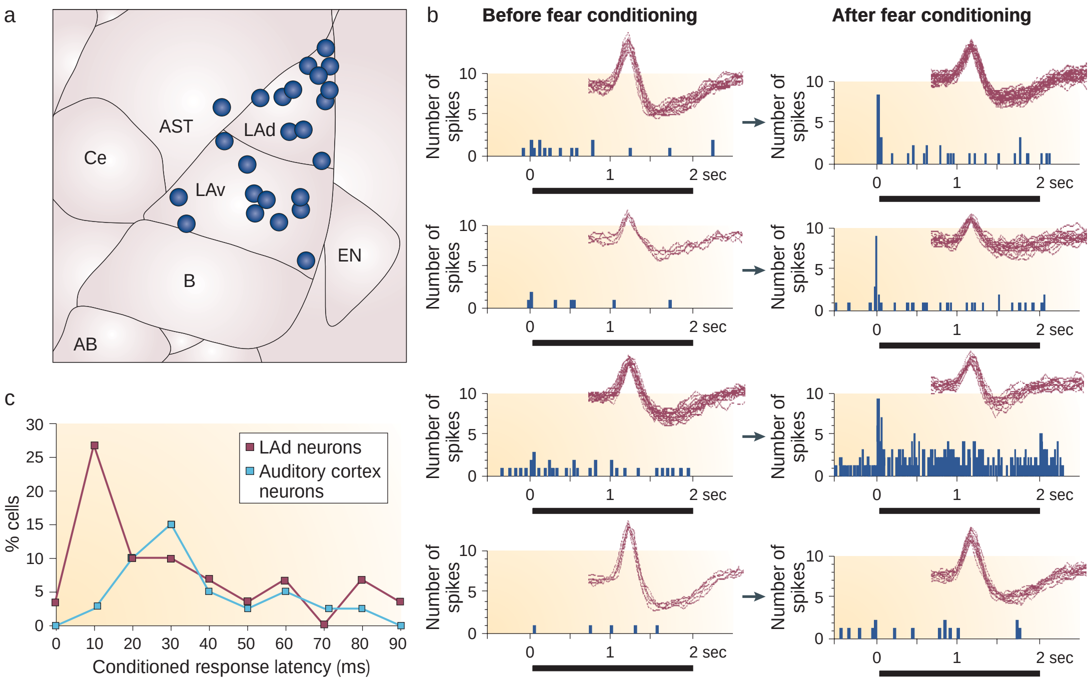
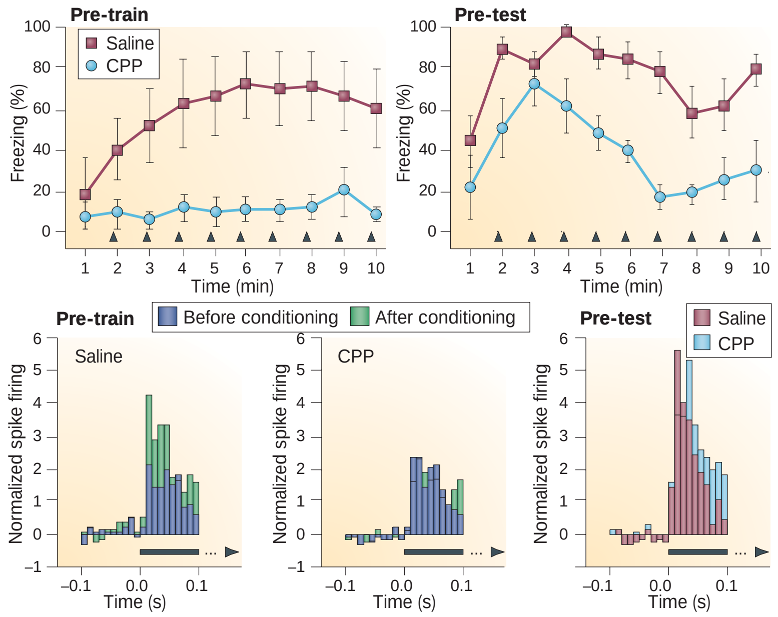


Figure 2 | **Effects of fear conditioning on lateral amygdala neurons.** Fear conditioning induces increases in conditional stimulus (CS)-evoked spike firing in lateral amygdala (LA) neurons. **a** | Electrode placements in the dorsal (LAd) and ventral (LAv) divisions of the lateral amygdala. AB, accessory basal nucleus; AST, amygdalo-striatal transition zone; B, basolateral nucleus; Ce, central nucleus of the amygdala; EN, endopiriform nucleus. **b** | Peri-event time histograms from eight simultaneously recorded single units in the LA. Each histogram represents the sum of ten CS presentations (black bar) before or after fear conditioning. Representative spike waveforms for each unit are shown as pink lines in the insets. **c** | Neurons in the LAd show conditioned increases in spike firing at shorter latencies (from CS onset) than do auditory cortical neurons. Adapted, with permission, from REF. 52 © (1997) Cell Press.

LONG-TERM POTENTIATION (LTP) An enduring increase in the amplitude of excitatory postsynaptic potentials as a result of high-frequency (tetanic) stimulation of afferent pathways. It is measured both as the amplitude of excitatory postsynaptic potentials and as the magnitude of the postsynaptic cell-population spike. LTP is most frequently studied in the hippocampus and is often considered to be the cellular basis of learning and memory in vertebrates.

Box 2 | NMDA receptors and associative plasticity in the amygdala



There is considerable evidence that long-term synaptic plasticity in the lateral amygdala (LA) mediates the acquisition of fear memory (see REFS 98–100 for reviews). There is strong evidence that the NMDA (*N*-methyl-D-aspartate) subclass of glutamate receptors is involved in both the acquisition of fear memory and the induction of long-term potentiation (LTP) in the amygdala[44,101], and although there is debate concerning the role of NMDA receptors in the expression of learned fear responses[102,103], recent work indicates that NMDA receptors might be selectively involved in fear-memory acquisition under some conditions[104]. A recent experiment by Maren and colleagues (see figure) examined whether NMDA receptors are also involved in the acquisition of associative neuronal activity in the LA during fear conditioning[105]. In this experiment, CPP (3-(2-carboxypiperazin-4-yl) propyl-1-phosphonic acid), a competitive NMDA-receptor antagonist, was administered either before training (pre-train) or before retention testing (pre-test) to examine the influence of NMDA-receptor blockade on the acquisition and expression, respectively, of conditional freezing and LA unit activity. Systemic administration of CPP impaired both the acquisition of auditory fear conditioning (as indexed by conditional freezing; arrowheads indicate conditional stimulus (CS) presentations) and conditioning-related increases in CS-elicited spike firing (pre-train panels; first 100 ms of the 2-second CS is indicated by the black bar and arrow). Although CPP completely eliminated the acquisition of conditional fear and associative spike firing in the LA, it had only a mild effect on the expression of these responses (pre-test panels). That is, CPP administered before a retention test in previously conditioned animals moderately attenuated conditional freezing, but did not reduce the magnitude of conditional spike firing in the LA. These data are consistent with models of fear conditioning that posit a role for NMDA-receptor-dependent synaptic plasticity in the formation of fear memory, and reveal that similar neurochemical mechanisms underlie the induction of amygdaloid LTP, conditioning-related increases in spike firing and conditional fear behaviour. Modified, with permission, from REF. 105 © (2004) Blackwell Publishing.

cortical area Te3 during auditory fear conditioning in rats[52]. Te3 is the auditory association area that projects to the LAd[53,54]. They observed that conditioned plasticity in Te3 neurons occurred later than in the LAd (30–50 ms versus 10–20 ms; FIG. 2c). Also, LAd neurons developed conditioned responses within the first three trials of fear conditioning, whereas Te3 neurons required between six and nine conditioning trials to show conditioned responses. Therefore, plasticity in the LAd is not likely to be fed forward passively from Te3, because it precedes Te3 both within and across trials.

It remains possible that LA plasticity is passively fed forward from the MGm/PIN. However, this seems unlikely, because inactivation of the BASOLATERAL AMYGDALA (BLA) with the $GABA_A$ (γ-aminobutyric acid, type A) receptor agonist muscimol prevents the acquisition of fear conditioning, as well as the expression of fear memory, 24 hours after training when rats are tested drug-free[55–57]. Therefore, the primary site of plasticity in fear conditioning is unlikely to be the MGm/PIN, although an effect of muscimol on brainstem projections that regulate ascending modulation of the thalamus cannot be ruled out.

An alternative explanation is that plasticity in thalamic or cortical neurons depends on the amygdala. To address this issue, Maren and colleagues used muscimol to inactivate the BLA while recording single-unit activity in the MGm/PIN[58]. In addition to preventing the development of conditioned fear, muscimol in the amygdala prevented the development of unit plasticity in the MGm/PIN. A similar observation was made for INSTRUMENTAL AVOIDANCE LEARNING in rabbits[59]. In a related experiment, Armony and co-workers recorded single-unit activity from cortical area Te3 in rats that had first received BLA lesions[60]. Although short-latency plastic responses were still observed in amygdala-lesioned rats, long-latency responses anticipating the onset of footshock were lost. Because muscimol inactivation of the BLA prevents the development of conditioned fear responses[57,58], amygdala-independent short-latency plasticity in Te3 does not seem to be sufficient to drive fear behaviour, and might represent associative learning at a more cognitive level[61]. By contrast, the loss of shock-anticipatory responses in Te3 neurons indicates that ascending projections from the amygdala might 'interrupt' cortical processing when danger is imminent[62].

Rather than mirroring thalamic or cortical plasticity, it seems that conditioning-related spike firing in the amygdala is independent of — and in some cases essential for — plasticity in the MGm/PIN and Te3. In fact, the LAd seems to be the first site in the auditory pathway to show associative plasticity that is not fed forward passively from upstream sites, is not dependent on downstream sites and is crucial for conditioned fear behaviour. Furthermore, LA neurons seem to drive plasticity at both thalamic and cortical levels, indicating that the amygdala facilitates memory storage in widespread areas, as shown by McGaugh and co-workers for inhibitory avoidance[63–65]. However, several important issues need to be resolved before we can conclude that the LA is a primary site of plasticity in fear conditioning, such as how LA spike firing relates to behaviour and the frequency specificity of LA plasticity in auditory fear conditioning (BOX 3).

Associative coding in the amygdala

For any conditioning-induced change in neuronal activity, it is essential to determine whether the change is related to the associative learning that encodes the CS–US contingency or whether it represents a non-associative process (a form of learning that does not depend on a CS–US association) that is consequent to

BASOLATERAL AMYGDALA
The region of the amygdala that encompasses the lateral, basolateral and basomedial nuclei.

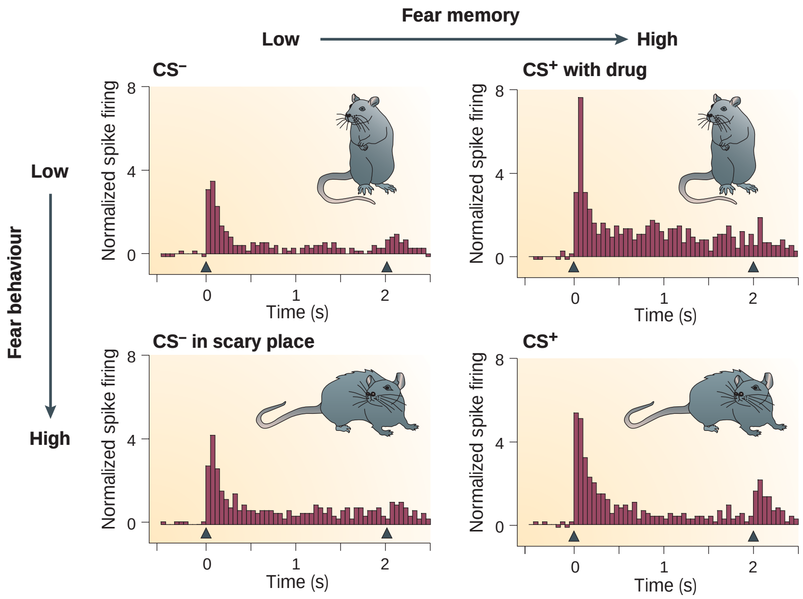


Figure 3 | **Lateral amygdala neurons encode fear memory independently of fear behaviour.** Each panel shows population averages for single units recorded in the lateral amygdala (LA) during presentations of an auditory cue paired with a footshock (CS$^+$) or an auditory cue that has never been paired with a shock (CS$^-$). Onset and offset of the auditory CSs are indicated by arrowheads. Fear conditioning increases both CS-evoked spike firing and freezing behaviour to the CS$^+$ (bottom right), but not to the CS$^-$ (top left). This typical correlation between the associative history of the CS and freezing behaviour can be broken by testing a CS$^-$ in a context that has been paired with unsignalled shock (CS$^-$ in scary place; bottom left) or by testing a CS$^+$ after inactivating the central nucleus of the amygdala (CS$^+$ after drug; top right). In these cases, the CS$^-$ is presented against a background of high fear behaviour, or the CS$^+$ is presented to animals that are not capable of showing conditioned fear responses. Nonetheless, LA neurons continue to show activity patterns that are consistent with the associative history of the CS$^-$ and CS$^+$; that is, LA neurons represent fear memory, and are not biased by the performance of fear responses. Adapted, with permission, from REF. 73 © (2003) Cell Press.

INSTRUMENTAL AVOIDANCE LEARNING
Instrumental learning is a form of learning that takes place through reinforcement (or punishment) that is contingent on the performance (or withholding) of a particular behaviour. So, the subject's response is instrumental in producing an outcome. Compare with Pavlovian learning.

EXTINCTION
The reduction in the conditioned response after non-reinforced presentations of the conditional stimulus.

RECEPTIVE FIELD
That limited domain of the sensory environment to which a given sensory neuron is responsive, such as a limited frequency band in audition or a limited area of space in vision.

either CS or US exposure. It is possible, for example, that increases in the responsiveness of LA neurons to auditory CSs are due to non-associative learning processes such as sensitization or pseudo-conditioning. Moreover, changes in behaviour and arousal that accompany learned fear might alter sensory processing in the brain in a way that mirrors associative learning but is not itself the substance of memory[6].

Quirk and colleagues[39] showed that CS-elicited firing in the LA was greater after CS–US pairings than with an earlier phase of unpaired CS and US presentations. This implies that LA firing is regulated by the associative contingency between the CS and the US. However, it is also possible that shock exposure during conditioning promoted further non-associative sensitization of spike firing to the CS. If so, changes in CS-evoked spike firing after conditioning might have resulted from nonspecific changes in the responsivity of amygdala neurons to any auditory stimulus, rather than an associative change to the specific CS paired with the US.

To assess this possibility, Paré and colleagues used a discriminative fear-conditioning procedure in conscious cats to determine the specificity of LA plasticity for the auditory CS paired with the US[66]. In this procedure, there were two distinct auditory cues: a CS$^+$ that was paired with a US, and a CS$^-$ that was not. In such a design, differential behaviour to the two CSs is taken as an index of associative learning, and changes in behaviour to the CS$^-$ relative to the pre-conditioning baseline are taken as an index of non-associative sensitization. Of course, the CSs must be chosen carefully to avoid generalization between the cues, which would mask the different associative strengths of the CSs.

Collins and Paré[66] found that discriminative fear conditioning produced CS-specific changes in fear behaviour, single units and local field potentials in the LA; that is, after fear conditioning, the CS$^+$ (a 5- or 10-kHz pure tone) evoked a larger LA field potential and more spike firing than it did before conditioning. Conversely, fear conditioning decreased the field potentials and spike firing that were elicited by the CS$^-$. These changes in CS-elicited neural activity also showed EXTINCTION, returning to baseline levels after several presentations of each CS without the US. Therefore, the increased spike firing in the LA after fear conditioning is CS-specific and cannot be explained by a nonspecific sensitization of spike firing to auditory stimuli or to pseudo-conditioning. It should be noted, however, that a complete frequency RECEPTIVE FIELD analysis[61] has not yet been carried out in the LA.

Conditioning-related changes in LA activity are closely correlated with the expression of fear responses. Presentations of CSs that have been paired with a footshock evoke behavioural responses, such as freezing or an increased state of arousal associated with fear[67–69]. In many cases, these fear responses outlast the stimuli that produce them, and might therefore affect the processing of subsequent CSs. For example, LA neurons in cats that have undergone auditory fear conditioning show increased responsiveness not only to the auditory CS, but also to electrical activation of cortical inputs[70]. Because the cortical stimulation was never explicitly paired with the shock US in these animals, the potentiation of these responses might reflect nonspecific increases in LA excitability. A similar change in the intrinsic excitability of LA neurons has been observed after olfactory conditioning in rats[71].

Therefore, it is necessary to determine whether associative plasticity of CS-elicited LA spike firing is a cause of learned fear responses or a consequence of the behavioural changes that are engendered by the fear state. One approach to this question is to examine the development of neuronal plasticity over the course of conditioning[12]. If LA firing codes for fear associations, learning-related activity in the LA should occur before (or coincident with) the emergence of fear CRs. Repa and colleagues addressed this question by examining spike firing in the LA during the gradual acquisition of CONDITIONED LEVER-PRESS SUPPRESSION[72]. Interestingly, most of the neurons that were recorded in the LA showed increases in CS-elicited spike firing on or before the trial in which the first significant behavioural CR appeared. There were also neurons that increased their firing to the CS after this point. Moreover, some LA neurons maintained their conditioning-related increase in spike firing after extinction of the fear response, indicating that the expression of fear behaviour is not driving LA responsiveness.

Box 3 | Localizing fear memory

Fear conditioning increases the responses of single lateral amygdala (LA) neurons to the conditional stimulus (CS). However, this observation alone is not sufficient to imply that LA neurons signal fear memory. Additional criteria (all of which are met by the LA) are as follows:

Is plasticity in the LA associative?

Yes. LA neurons increase their tone responses during conditioning in contrast to pseudoconditioning (unpaired tones and shocks). Increases are specific to stimuli that are paired with a shock (CS^+), and are not seen with unpaired stimuli ($CS^–$).

Does plasticity in the LA depend on plasticity in the auditory cortex?

No. Plasticity in the LA precedes plasticity in the auditory cortex, both within and across training trials.

Does plasticity in the LA depend on plasticity in the auditory thalamus?

Probably not. Inactivation of the LA with the $GABA_A$ (γ-aminobutyric acid, type A) agonist muscimol prevents the development of plasticity in medial geniculate inputs to the LA. Therefore, plasticity in the medial geniculate nucleus seems to depend on plasticity in the LA.

Do LA neurons learn as fast as the rat learns?

Yes. Across trials, plasticity in the LA develops as fast as — or faster than — conditioned fear responses.

Is plasticity in the LA caused by fear behaviour?

No. Plasticity in LA neurons can be dissociated from freezing behaviour, implying that LA neurons signal the strength of the conditional–unconditional stimulus association rather than fear *per se*.

In a more direct examination of this issue, Goosens and colleagues recently asked whether increases in LA spike firing are caused by the expression of conditional freezing behaviour[73] (FIG. 3). In one experiment, rats received discriminative fear conditioning using distinct auditory CSs. Separate groups of animals were then tested to each CS in either a neutral context (control group) or in a context that they had come to fear through contextual fear conditioning (experimental group). In this way, it could be determined whether fear *per se* was sufficient to alter LA spike firing to a cue ($CS^–$) that was not paired with a footshock. In fact, the expression of fear behaviour did not alter LA spike firing, and the degree of neuronal discrimination between the control and experimental rats was nearly identical. In a follow-up experiment, the influence of inhibiting the expression of conditional freezing on LA plasticity was explored[72]. Reversible inhibition of the central nucleus of the amygdala eliminated conditional freezing behaviour but not associative increases in CS-elicited spike firing in the LA.

Together, these experiments show that the expression of fear is neither sufficient nor necessary for the expression of associative plasticity in the LA, supporting the view that LA neurons encode fear memories. The essence of this mnemonic code seems to be contained in the rate at which LA neurons fire action potentials in response to auditory CSs. In addition to this rate code, however, the LA might also signal fear associations by the timing of spikes within a CS-evoked spike train: a rhythm code. Fear conditioning has been shown to increase synchrony in LA neurons[39,70], and THETA OSCILLATIONS become more frequent in the LA after fear conditioning[70,74]. It has been suggested that increased synchrony after fear conditioning could increase the impact of the LA on neocortical targets that consolidate and store emotional memories[75].

Fear not: amygdala inhibition after extinction

Fear memories enable us to anticipate and respond to dangers in our environments. However, when signals for aversive events no longer predict those events, fear to those signals subsides. This inhibitory learning process, known as extinction, has important clinical relevance as a treatment for anxiety disorders, such as panic disorder[76] and post-traumatic stress[77]. Importantly, the inhibitory memories that are learned during extinction compete with the excitatory memories that are formed during conditioning, thereby suppressing fear responses[78]. Although fear subsides after extinction, the fear memory is not erased. In fact, the inhibitory memories of extinction are relatively short-lived and context-dependent. This means that extinction is expressed only in the context in which extinction was given, and even in that context, fear responses will spontaneously recover over time[79]. This transience and context dependence of extinction implies that biology has deemed it better to fear than not to fear.

There is considerable interest in understanding the neurobiological mechanisms of fear extinction, and substantial progress has been made in recent years[80,81]. As for fear conditioning, the amygdala seems to have a vital role in the extinction of learned fear. Pharmacological manipulations that inhibit neuronal activity or disrupt the cellular processes that underlie synaptic plasticity in the amygdala impair extinction[82,83]. The mediation of extinction by the amygdala is also manifested in the firing of LA neurons. Presenting the CS in the absence of the US reduces the expression of both behavioural CRs and CS-evoked spike firing in most LA neurons[39,72]. However, not all LA neurons reduce their firing after extinction[72], and even neurons that do reduce their firing continue to show the synchrony that is fostered by conditioning[39]. This implies that even after extinction, residual traces of conditioning persist in the activity patterns of LA neurons.

The reduction in CS-evoked spike firing in the LA that accompanies extinction correlates with the attenuation of fear CRs to the extinguished CS. However, as described earlier, fear extinction is context-dependent and is primarily expressed only in the extinction context. This raises the question of whether the suppression in LA spike firing after extinction is also context-dependent. To address this question, Hobin and colleagues used an elegant within-subjects behavioural design to observe the activity that is elicited in LA neurons by extinguished CSs that are presented either within or outside their extinction context[84]. Rats were conditioned to fear two distinct auditory CSs, then they received extinction training to each CS in a different context. Neurophysiological recordings were taken in a series of four test sessions, in which each CS was tested in each context. This design eliminated the possibility that any particular CS, context or CS/context combination

CONDITIONED LEVER-PRESS SUPPRESSION
The reduction in pressing for food reward in the presence of a fear-conditioned stimulus.

THETA OSCILLATIONS
Rhythmic neural activity with a frequency of 4–8 Hz.

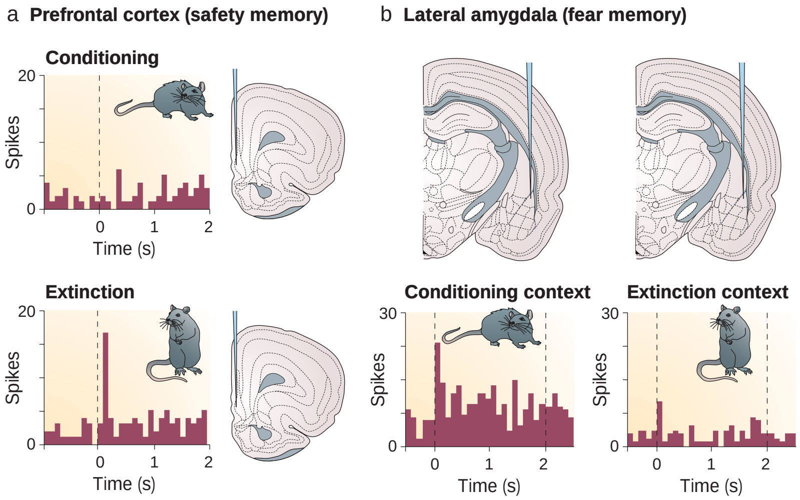


Figure 4 | **Neuronal signalling of extinction in the prefrontal cortex and lateral amygdala.** Panels show a representative single unit recorded from the infralimbic region of the medial prefrontal cortex (PFC; **a**) and the lateral amygdala (LA; **b**). **a** | Unlike neurons in the LA, PFC neurons are initially silent during conditional stimulus (CS) presentations after fear conditioning (conditioning), but greatly increase their CS-elicited firing after extinction training (extinction). **b** | Although spike firing is inhibited in the LA by extinction training (extinction context), it can be renewed by a change in context (conditioning context). These data reveal that neurons in both the PFC and LA respond to extinction contingencies, although they respond in opposite directions under these conditions. Adapted, with permission, from REF. 84 © (2003) Society for Neuroscience, and from REF. 88 © (2002) Macmillan Magazines Ltd.

PREFRONTAL CORTEX
(PFC) The non-motor sectors of the frontal lobe that receive input from the dorsomedial thalamic nucleus and subserve working memory, complex attentional processes and executive functions such as planning, behavioural inhibition, logical reasoning, action monitoring and social cognition.

might itself affect LA spike firing independently of the extinction history of the CS and context. Interestingly, most single units in the LA modulated their firing rates to extinguished CSs according to the context in which the CS was presented. When a CS was presented in the extinction context, spike firing to that CS was typically lower than when the CS was presented outside its extinction context; a small number of neurons showed the opposite pattern of modulation. However, the population average mirrored the behavioural expression of fear, indicating that the context dependence of extinguished fear is modulated at the level of the LA (FIG. 4).

It is of considerable interest to understand how LA activity and fear expression are modulated after extinction. Recent data indicate an important role for the medial PREFRONTAL CORTEX (mPFC). Rats with mPFC lesions can learn to extinguish fear CRs, but have difficulty recalling the extinction memory 24 hours after training[85–87]. This is precisely the time when mPFC neurons show robust increases in CS-elicited firing[88,89], consistent with a role in inhibition of fear after extinction (FIG. 4). mPFC neurons show an inhibitory influence on both the LA[90] and the central nucleus[91], the main output regions of the amygdala. Furthermore, pairing CSs with electrical stimulation of the mPFC mimics extinction behaviour[88,92]. Electrical stimulation of the mPFC inhibits both lateral and central amygdaloid neurons, presumably through a rich network of inhibitory interneurons embedded in the amygdala[93,94] (FIG. 5).

If the inhibitory signal for extinction originates in the mPFC, then it is probably modulated by context. One possible modulator of the mPFC is the hippocampus. A recent study indicates that the hippocampus modulates the expression of extinction memories[95]. Temporary inactivation of the dorsal hippocampus with muscimol eliminated renewal of fear to an extinguished CS; extinction performance prevailed under conditions in which it would normally be weak. This implies that although the hippocampus is not the repository for extinction memories, it is involved in regulating when and where extinction memories are expressed. The mechanism by which the hippocampus interacts with the amygdala to regulate CS-evoked spike firing is not clear, and could involve either a direct projection from the hippocampal formation to the LA[44,96] or an indirect projection through the prefrontal cortex[97] (FIG. 5).

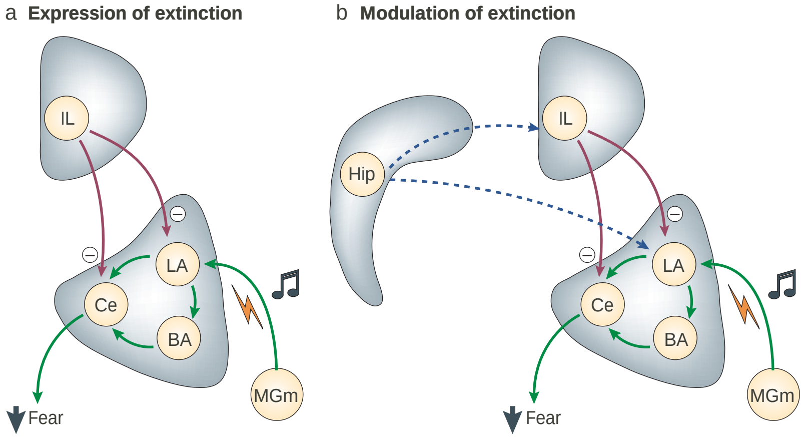


Figure 5 | **Cortical modulation of amygdala fear memories in extinction. a** | Following extinction, neurons in the infralimbic region of the medial prefrontal cortex (IL) increase their responses to tones. The IL exerts feed-forward inhibition of neurons in the lateral amygdala (LA) and the central nucleus of the amygdala (Ce), thereby decreasing the expression of fear memories. **b** | Extinction is expressed only in the context in which it occurred. Contextual modulation of extinction requires the involvement of the hippocampus (Hip), which could modulate fear responses either at the level of the LA or the IL. BA, basal amygdala; MGm, medial subdivision of the medial geniculate nucleus.

Conclusions

Numerous studies have revealed electrophysiological correlates of memory in neuronal activity patterns of behaving animals, but few of these studies have established causality between learning-induced changes in neuronal activity and behaviour. Recent work in fear conditioning renews the promise of localizing memory in neuronal activity patterns in the mammalian brain. LA neurons seem to be the origin of associative plasticity that is relevant for both learned behavioural responses and physiological plasticity in other brain regions after aversive conditioning. Moreover, modulation of the fear-memory code in the LA is involved in the suppression and renewal of fear responses after extinction.

This research opens up new avenues to investigate how the hippocampus, prefrontal cortex and amygdala interact during the acquisition, storage and retrieval of fear memories, and how cellular and synaptic mechanisms encode inhibitory extinction memories together with excitatory fear memories. The central role for amygdala neurons in both processes reveals a common target for clinical interventions for anxiety disorders.

1. Hebb, D. O. *The Organization of Behavior.* (John Wiley and Sons, New York, 1949).
2. Nicolelis, M. A. & Ribeiro, S. Multielectrode recordings: the next steps. *Curr. Opin. Neurobiol.* **12**, 602–606 (2002).
3. Buzsaki, G. Large-scale recording of neuronal ensembles. *Nature Neurosci.* **7**, 446–451 (2004).
4. Galambos, R., Myers, R. & Sheatz, G. Extralemniscal activation of auditory cortex in cats. *Am. J. Physiol.* **200**, 23–28 (1961).
5. Gerken, G. M. & Neff, W. D. Experimental procedures affecting evoked responses recorded from auditory cortex. *Electroencephalogr. Clin. Neurophysiol.* **15**, 947–957 (1963).
6. Hall, R. D. & Mark, R. G. Fear and the modification of acoustically evoked potentials during conditioning. *J. Neurophysiol.* **30**, 893–910 (1967).
7. Kamikawa, K., Mcilwain, J. T. & Adey, W. R. Response of thalamic neurons during classical conditioning. *Electroencephalogr. Clin. Neurophysiol.* **17**, 485–496 (1964).
8. O'Brien, J. H. & Fox, S. S. Single-cell activity in cat motor cortex. I. Modifications during classical conditioning procedures. *J. Neurophysiol.* **32**, 267–284 (1969).
9. Woody, C. D., Vassilevsky, N. N. & Engel, J. Conditioned eye blink: unit activity at coronal-precruciate cortex of the cat. *J. Neurophysiol.* **33**, 851–864 (1970).
10. Oleson, T. D., Ashe, J. H. & Weinberger, N. M. Modification of auditory and somatosensory system activity during pupillary conditioning in the paralyzed cat. *J. Neurophysiol.* **38**, 1114–1139 (1975).
11. Weinberger, N. M., Imig, T. J. & Lippe, W. R. Modification of unit discharges in the medial geniculate nucleus by click-shock pairing. *Exp. Neurol.* **36**, 46–58 (1972).
12. Olds, J., Disterhoft, J. F., Segal, M., Kornblith, C. L. & Hirsh, R. Learning centers of rat brain mapped by measuring latencies of conditioned unit responses. *J. Neurophysiol.* **35**, 202–219 (1972).
A landmark study that describes a methodology for using single-unit response latencies to auditory stimuli to localize sites of neuronal plasticity in the brain during learning.
13. Gabriel, M. Short-latency discriminative unit response: Engram or bias? *Physiol. Psychol.* **4**, 275–280 (1976).
14. Disterhoft, J. F. & Stuart, D. K. Trial sequence of changed unit activity in auditory system of alert rat during conditioned response acquisition and extinction. *J. Neurophysiol.* **39**, 266–281 (1976).
15. Disterhoft, J. F. & Olds, J. Differential development of conditioned unit changes in thalamus and cortex of rat. *J. Neurophysiol.* **35**, 665–679 (1972).
16. Medina, J. F., Christopher, R. J., Mauk, M. D. & LeDoux, J. E. Parallels between cerebellum- and amygdala-dependent conditioning. *Nature Rev. Neurosci.* **3**, 122–131 (2002).
17. Christian, K. M. & Thompson, R. F. Neural substrates of eyeblink conditioning: acquisition and retention. *Learn. Mem.* **10**, 427–455 (2003).
18. Ben Ari, Y. & Le Gal la Salle, G. Plasticity at unitary level. II. Modifications during sensory–sensory association procedures. *Electroencephalogr. Clin. Neurophysiol.* **32**, 667–679 (1972).
19. McGaugh, J. L. Hormonal influences on memory. *Annu. Rev. Psychol.* **34**, 297–323 (1983).
20. Kapp, B. S., Frysinger, R. C., Gallagher, M. & Haselton, J. R. Amygdala central nucleus lesions: effect on heart rate conditioning in the rabbit. *Physiol. Behav.* **23**, 1109–1117 (1979).
One of the earliest reports to describe a disruption of Pavlovian fear conditioning after selective amygdala lesions, indicating that the amygdala might be a site of plasticity in fear learning.
21. Krettek, J. E. & Price, J. L. A description of the amygdaloid complex in the rat and cat with observations on intra-amygdaloid axonal connections. *J. Comp. Neurol.* **178**, 255–280 (1978).
22. Hopkins, D. A. & Holstege, G. Amygdaloid projections to the mesencephalon, pons and medulla oblongata in the cat. *Exp. Brain Res.* **32**, 529–547 (1978).
23. Applegate, C. D., Frysinger, R. C., Kapp, B. S. & Gallagher, M. Multiple unit activity recorded from amygdala central nucleus during Pavlovian heart rate conditioning in rabbit. *Brain Res.* **238**, 457–462 (1982).
24. Pascoe, J. P. & Kapp, B. S. Electrophysiological characteristics of amygdaloid central nucleus neurons during Pavlovian fear conditioning in the rabbit. *Behav. Brain Res.* **16**, 117–133 (1985).
25. Iwata, J., LeDoux, J. E., Meeley, M. P., Arneric, S. & Reis, D. J. Intrinsic neurons in the amygdaloid field projected to by the medial geniculate body mediate emotional responses conditioned to acoustic stimuli. *Brain Res.* **383**, 195–214 (1986).
26. LeDoux, J. E., Sakaguchi, A. & Reis, D. J. Subcortical efferent projections of the medial geniculate nucleus mediate emotional responses conditioned to acoustic stimuli. *J. Neurosci.* **4**, 683–698 (1984).
27. LeDoux, J. E., Sakaguchi, A., Iwata, J. & Reis, D. J. Interruption of projections from the medial geniculate body to an archi-neostriatal field disrupts the classical conditioning of emotional responses to acoustic stimuli. *Neuroscience* **17**, 615–627 (1986).
28. Pitkanen, A., Savander, V. & LeDoux, J. E. Organization of intra-amygdaloid circuitries in the rat: an emerging framework for understanding functions of the amygdala. *Trends Neurosci.* **20**, 517–523 (1997).
29. Paré, D. & Smith, Y. Intrinsic circuitry of the amygdaloid complex: common principles of organization in rats and cats. *Trends Neurosci.* **21**, 240–241 (1998).
30. LeDoux, J. E., Farb, C. & Ruggiero, D. A. Topographic organization of neurons in the acoustic thalamus that project to the amygdala. *J. Neurosci.* **10**, 1043–1054 (1990).
31. LeDoux, J. E., Ruggiero, D. A. & Reis, D. J. Projections to the subcortical forebrain from anatomically defined regions of the medial geniculate body in the rat. *J. Comp. Neurol.* **242**, 182–213 (1985).
32. LeDoux, J. E., Cicchetti, P., Xagoraris, A. & Romanski, L. M. The lateral amygdaloid nucleus: sensory interface of the amygdala in fear conditioning. *J. Neurosci.* **10**, 1062–1069 (1990).
33. Romanski, L. M. & LeDoux, J. E. Equipotentiality of thalamo–amygdala and thalamo–cortico–amygdala circuits in auditory fear conditioning. *J. Neurosci.* **12**, 4501–4509 (1992).
34. Fanselow, M. S. & LeDoux, J. E. Why we think plasticity underlying Pavlovian fear conditioning occurs in the basolateral amygdala. *Neuron* **23**, 229–232 (1999).
35. Vazdarjanova, A. & McGaugh, J. L. Basolateral amygdala is not critical for cognitive memory of contextual fear conditioning. *Proc. Natl Acad. Sci. USA* **95**, 15003–15007 (1998).
36. Killcross, A. S., Robbins, T. W. & Everitt, B. J. Different types of fear-conditioned behavior mediated by separate nuclei within the amygdala. *Nature* **388**, 377–380 (1997).
37. Uwano, T., Nishijo, H., Ono, T. & Tamura, R. Neuronal responsiveness to various sensory stimuli, and associative learning in the rat amygdala. *Neuroscience* **68**, 339–361 (1995).
38. Ben Ari, Y. & Le Gal la Salle, G. Lateral amygdala unit activity: II. Habituating and non-habituating neurons. *Electroencephalogr. Clin. Neurophysiol.* **37**, 463–472 (1974).
39. Quirk, G. J., Repa, C. & LeDoux, J. E. Fear conditioning enhances short-latency auditory responses of lateral amygdala neurons: parallel recordings in the freely behaving rat. *Neuron* **15**, 1029–1039 (1995).
This study was the first to use multiple single-unit recordings to describe short-latency plasticity in LA neurons, consistent with potentiation of inputs from the auditory thalamus during fear conditioning.
40. Li, X. F., Stutzmann, G. E. & LeDoux, J. E. Convergent but temporally separated inputs to lateral amygdala neurons from the auditory thalamus and auditory cortex use different postsynaptic receptors: *in vivo* intracellular and extracellular recordings in fear conditioning pathways. *Learn. Mem.* **3**, 229–242 (1996).
41. Maren, S. Auditory fear conditioning increases CS-elicited spike firing in lateral amygdala neurons even after extensive overtraining. *Eur. J. Neurosci.* **12**, 4047–4054 (2000).
42. Clugnet, M. C. & LeDoux, J. E. Synaptic plasticity in fear conditioning circuits: induction of LTP in the lateral nucleus of the amygdala by stimulation of the medial geniculate body. *J. Neurosci.* **10**, 2818–2824 (1990).
43. Chapman, P. F., Kairiss, E. W., Keenan, C. L. & Brown, T. H. Long-term synaptic potentiation in the amygdala. *Synapse* **6**, 271–278 (1990).
A seminal paper that demonstrated for the first time that amygdala neurons show long-term synaptic potentiation *in vitro*.
44. Maren, S. & Fanselow, M. S. Synaptic plasticity in the basolateral amygdala induced by hippocampal formation stimulation *in vivo*. *J. Neurosci.* **15**, 7548–7564 (1995).
45. Rogan, M. T., Staubli, U. V. & LeDoux, J. E. Fear conditioning induces associative long-term potentiation in the amygdala. *Nature* **390**, 604–607 (1997).
An important paper showing that the acquisition of conditional fear responses is associated with physiological changes in auditory-evoked potentials in the amygdala, consistent with the induction of LTP.
46. McKernan, M. G. & Shinnick-Gallagher, P. Fear conditioning induces a lasting potentiation of synaptic currents *in vitro*. *Nature* **390**, 607–611 (1997).
47. Tsvetkov, E., Carlezon, W. A., Benes, F. M., Kandel, E. R. & Bolshakov, V. Y. Fear conditioning occludes LTP-induced presynaptic enhancement of synaptic transmission in the cortical pathway to the lateral amygdala. *Neuron* **34**, 289–300 (2002).
An elegant study using behavioural and *in vitro* electrophysiological techniques to show that training occludes synaptic increases in presynaptic neurotransmitter release after LTP induction in LA neurons. This provides strong evidence that fear conditioning is mediated by LTP in the amygdala.
48. Bordi, F. & LeDoux, J. E. Response properties of single units in areas of rat auditory thalamus that project to the amygdala. II. Cells receiving convergent auditory and somatosensory inputs and cells antidromically activated by amygdala stimulation. *Exp. Brain Res.* **98**, 275–286 (1994).
49. Romanski, L. M., Clugnet, M. C., Bordi, F. & LeDoux, J. E. Somatosensory and auditory convergence in the lateral nucleus of the amygdala. *Behav. Neurosci.* **107**, 444–450 (1993).
50. Edeline, J. M. & Weinberger, N. M. Associative retuning in the thalamic source of input to the amygdala and auditory cortex: receptive field plasticity in the medial division of the medial geniculate body. *Behav. Neurosci.* **106**, 81–105 (1992).
51. Edeline, J. M., Neuenschwander-el Massioui, N. & Dutrieux, G. Discriminative long-term retention of rapidly induced multiunit changes in the hippocampus, medial geniculate and auditory cortex. *Behav. Brain Res.* **39**, 145–155 (1990).
52. Quirk, G. J., Armony, J. L. & LeDoux, J. E. Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. *Neuron* **19**, 613–624 (1997).
53. LeDoux, J. E., Farb, C. R. & Romanski, L. M. Overlapping projections to the amygdala and striatum from auditory processing areas of the thalamus and cortex. *Neurosci. Lett.* **134**, 139–144 (1991).
54. Romanski, L. M. & LeDoux, J. E. Information cascade from primary auditory cortex to the amygdala: corticocortical and corticoamygdaloid projections of temporal cortex in the rat. *Cereb. Cortex* **3**, 515–532 (1993).
55. Helmstetter, F. J. & Bellgowan, P. S. Effects of muscimol applied to the basolateral amygdala on acquisition and expression of contextual fear conditioning in rats. *Behav. Neurosci.* **108**, 1005–1009 (1994).
56. Muller, J., Corodimas, K. P., Fridel, Z. & LeDoux, J. E. Functional inactivation of the lateral and basal nuclei of the amygdala by muscimol infusion prevents fear conditioning to an explicit conditioned stimulus and to contextual stimuli. *Behav. Neurosci.* **111**, 683–691 (1997).
57. Wilensky, A. E., Schafe, G. E. & LeDoux, J. E. Functional inactivation of the amygdala before but not after auditory fear conditioning prevents memory formation. *J. Neurosci.* **19**, RC48 (1999).
58. Maren, S., Yap, S. A. & Goosens, K. A. The amygdala is essential for the development of neuronal plasticity in the medial geniculate nucleus during auditory fear conditioning in rats. *J. Neurosci.* **21**, RC135 (2001).
59. Poremba, A. & Gabriel, M. Amygdalar efferents initiate auditory thalamic discriminative training-induced neuronal activity. *J. Neurosci.* **21**, 270–278 (2001).
60. Armony, J. L., Quirk, G. J. & LeDoux, J. E. Differential effects of amygdala lesions on early and late plastic components of auditory cortex spike trains during fear conditioning. *J. Neurosci.* **18**, 2592–2601 (1998).
61. Weinberger, N. M. Specific long-term memory traces in primary auditory cortex. *Nature Rev. Neurosci.* **5**, 279–290 (2004).
62. Armony, J. L. & LeDoux, J. E. How the brain processes emotional information. *Ann. NY Acad. Sci.* **821**, 259–270 (1997).
63. Roozendaal, B., McReynolds, J. R. & McGaugh, J. L. The basolateral amygdala interacts with the medial prefrontal cortex in regulating glucocorticoid effects on working memory impairment. *J. Neurosci.* **24**, 1385–1392 (2004).
64. McGaugh, J. L. The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annu. Rev. Neurosci.* **27**, 1–28 (2004).
65. Cahill, L. Neurobiological mechanisms of emotionally influenced, long-term memory. *Prog. Brain Res.* **126**, 29–37 (2000).
66. Collins, D. R. & Paré, D. Differential fear conditioning induces reciprocal changes in the sensory responses of lateral amygdala neurons to the CS^+ and CS^-. *Learn. Mem.* **7**, 97–103 (2000).
67. LeDoux, J. E. Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
68. Davis, M. in *The Amygdala* (ed. Aggleton, J. P.) 213–288 (Oxford Univ. Press, Oxford, 2000).
69. Maren, S. Neurobiology of Pavlovian fear conditioning. *Annu. Rev. Neurosci.* **24**, 897–931 (2001).
70. Paré, D. & Collins, D. R. Neuronal correlates of fear in the lateral amygdala: multiple extracellular recordings in conscious cats. *J. Neurosci.* **20**, 2701–2710 (2000).
71. Rosenkranz, J. A. & Grace, A. A. Dopamine-mediated modulation of odour-evoked amygdala potentials during Pavlovian conditioning. *Nature* **417**, 282–287 (2002).

This study was the first to use intracellular recording methods to show that fear conditioning increases the excitability of LA neurons.

72. Repa, J. C. *et al.* Two different lateral amygdala cell populations contribute to the initiation and storage of memory. *Nature Neurosci.* **4**, 724–731 (2001).
73. Goosens, K. A., Hobin, J. A. & Maren, S. Auditory-evoked spike firing in the lateral amygdala and Pavlovian fear conditioning: mnemonic code or fear bias? *Neuron* **40**, 1013–1022 (2003).
This is an important paper that shows that conditioning-related changes in CS-evoked single-unit activity in the LA can be dissociated from fear behaviour, providing support for a role for the amygdala in coding fear memories.
74. Seidenbecher, T., Laxmi, T. R., Stork, O. & Pape, H. C. Amygdalar and hippocampal theta rhythm synchronization during fear memory retrieval. *Science* **301**, 846–850 (2003).
75. Pelletier, J. G. & Paré, D. Role of amygdala oscillations in the consolidation of emotional memories. *Biol. Psychiatry* **55**, 559–562 (2004).
76. Bouton, M. E., Mineka, S. & Barlow, D. H. A modern learning theory perspective on the etiology of panic disorder. *Psychol. Rev.* **108**, 4–32 (2001).
77. Rothbaum, B. O. & Schwartz, A. C. Exposure therapy for posttraumatic stress disorder. *Am. J. Psychother.* **56**, 59–75 (2002).
78. Bouton, M. E., Rosengard, C., Achenbach, G. G., Peck, C. A. & Brooks, D. C. Effects of contextual conditioning and unconditional stimulus presentation on performance in appetitive conditioning. *Q. J. Exp. Psychol.* **46**, 63–95 (1993).
79. Quirk, G. J. Memory for extinction of conditioned fear is long-lasting and persists following spontaneous recovery. *Learn. Mem.* **9**, 402–407 (2002).
80. Myers, K. M. & Davis, M. Behavioral and neural analysis of extinction. *Neuron* **36**, 567–584 (2002).
81. Quirk, G. J. Learning not to fear, faster. *Learn. Mem.* **11**, 125–126 (2004).
82. Falls, W. A., Miserendino, M. J. & Davis, M. Extinction of fear-potentiated startle: blockade by infusion of an NMDA antagonist into the amygdala. *J. Neurosci.* **12**, 854–863 (1992).
83. Lu, K. T., Walker, D. L. & Davis, M. Mitogen-activated protein kinase cascade in the basolateral nucleus of amygdala is involved in extinction of fear-potentiated startle. *J. Neurosci.* **21**, RC162 (2001).
84. Hobin, J. A., Goosens, K. A. & Maren, S. Context-dependent neuronal activity in the lateral amygdala represents fear memories after extinction. *J. Neurosci.* **23**, 8410–8416 (2003).
85. Quirk, G. J., Russo, G. K., Barron, J. L. & Lebron, K. The role of ventromedial prefrontal cortex in the recovery of extinguished fear. *J. Neurosci.* **20**, 6225–6231 (2000).
86. Lebron, K., Milad, M. R., & Quirk, G. J. Delayed recall of fear extinction in rats with lesions of ventral medial prefrontal cortex. *Learn. Mem.* **11**, 544–548 (2004).
87. Morgan, M. A., Romanski, L. M. & LeDoux, J. E. Extinction of emotional learning: contribution of medial prefrontal cortex. *Neurosci. Lett.* **163**, 109–113 (1993).
88. Milad, M. R. & Quirk, G. J. Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature* **420**, 70–74 (2002).
This study provides neurophysiological support for Pavlov's hypothesis that extinction involves inhibition, by showing that extinction increases the firing rate of prefrontal cortical neurons, and electrical stimulation of the prefrontal cortex inhibits fear responses.
89. Herry, C. & Garcia, R. Prefrontal cortex long-term potentiation, but not long-term depression, is associated with the maintenance of extinction of learned fear in mice. *J. Neurosci.* **22**, 577–583 (2002).
90. Rosenkranz, J. A., Moore, H. & Grace, A. A. The prefrontal cortex regulates lateral amygdala neuronal plasticity and responses to previously conditioned stimuli. *J. Neurosci.* **23**, 11054–11064 (2003).
91. Quirk, G. J., Likhtik, E., Pelletier, J. G. & Paré, D. Stimulation of medial prefrontal cortex decreases the responsiveness of central amygdala output neurons. *J. Neurosci.* **23**, 8800–8807 (2003).
92. Milad, M. R., Vidal-Gonzalez, I. & Quirk, G. J. Electrical stimulation of medial prefrontal cortex reduces conditioned fear in a temporally specific manner. *Behav. Neurosci.* **118**, 389–395 (2004).
93. Royer, S., Martina, M. & Paré, D. An inhibitory interface gates impulse traffic between the input and output stations of the amygdala. *J. Neurosci.* **19**, 10575–10583 (1999).
This study showed that amygdala output could be inhibited by GABA-releasing intercalated neurons, implying that there is complex processing of fear signals within the amygdala. The inhibition of amygdala output by this mechanism might be important for fear extinction.
94. Szinyei, C., Heinbockel, T., Montagne, J. & Pape, H. C. Putative cortical and thalamic inputs elicit convergent excitation in a population of GABAergic interneurons of the lateral amygdala. *J. Neurosci.* **20**, 8909–8915 (2000).
95. Corcoran, K. A. & Maren, S. Hippocampal inactivation disrupts contextual retrieval of fear memory after extinction. *J. Neurosci.* **21**, 1720–1726 (2001).
96. Pitkanen, A., Pikkarainen, M., Nurminen, N. & Ylinen, A. Reciprocal connections between the amygdala and the hippocampal formation, perirhinal cortex, and postrhinal cortex in rat. A review. *Ann. NY Acad. Sci.* **911**, 369–391 (2000).
97. Thierry, A. M., Gioanni, Y., Degenetais, E. & Glowinski, J. Hippocampo–prefrontal cortex pathway: anatomical and electrophysiological characteristics. *Hippocampus* **10**, 411–419 (2000).
98. Maren, S. Long-term potentiation in the amygdala: a mechanism for emotional learning and memory. *Trends Neurosci.* **22**, 561–567 (1999).
99. Blair, H. T., Schafe, G. E., Bauer, E. P., Rodrigues, S. M. & LeDoux, J. E. Synaptic plasticity in the lateral amygdala: a cellular hypothesis of fear conditioning. *Learn. Mem.* **8**, 229–242 (2001).
An excellent review covering the cellular and synaptic mechanisms in the lateral amygdala that underlie the acquisition of long-term fear memories.
100. Schafe, G. E., Nader, K., Blair, H. T. & LeDoux, J. E. Memory consolidation of Pavlovian fear conditioning: a cellular and molecular perspective. *Trends Neurosci.* **24**, 540–546 (2001).
101. Miserendino, M. J., Sananes, C. B., Melia, K. R. & Davis, M. Blocking of acquisition but not expression of conditioned fear-potentiated startle by NMDA antagonists in the amygdala. *Nature* **345**, 716–718 (1990).
This is the first report to reveal a crucial role for amygdala NMDA receptors in the acquisition of Pavlovian fear conditioning.
102. Maren, S., Aharonov, G., Stote, D. L. & Fanselow, M. S. *N*-methyl-D-aspartate receptors in the basolateral amygdala are required for both acquisition and expression of conditional fear in rats. *Behav. Neurosci.* **110**, 1365–1374 (1996).
103. Fendt, M. Injections of the NMDA receptor antagonist aminophosphonopentanoic acid into the lateral nucleus of the amygdala block the expression of fear-potentiated startle and freezing. *J. Neurosci.* **21**, 4111–4115 (2001).
104. Rodrigues, S. M., Schafe, G. E. & LeDoux, J. E. Intra-amygdala blockade of the NR2B subunit of the NMDA receptor disrupts the acquisition but not the expression of fear conditioning. *J. Neurosci.* **21**, 6889–6896 (2001).
105. Goosens, K. A. & Maren, S. NMDA receptors are essential for the acquisition, but not expression, of conditional fear and associative spike firing in the lateral amygdala. *Eur. J. Neurosci.* **20**, 537–548 (2004).

Acknowledgements

The authors thank K. Goosens and two anonymous reviewers for helpful comments on the manuscript. This work was supported by grants from the National Institute of Mental Health.

Competing interests statement

The authors declare no competing financial interests.

Online links

FURTHER INFORMATION

Encyclopedia of Life Sciences: http://www.els.net/
GABA$_A$ receptors | Long-term potentiation | Neural informaton processing | NMDA receptors

Maren's laboratory: http://marenlab.org

Quirk's laboratory: http://www.psm.edu/Quirk%20Lab/index.htm

Access to this interactive links box is free online.

Annu. Rev. Psychol. 2006. 57:87–115
doi: 10.1146/annurev.psych.56.091103.070229
Copyright © 2006 by Annual Reviews. All rights reserved
First published online as a Review in Advance on September 16, 2005

BEHAVIORAL THEORIES AND THE NEUROPHYSIOLOGY OF REWARD

Wolfram Schultz
Department of Anatomy, University of Cambridge, CB2 3DY United Kingdom; email: ws234@cam.ac.uk

Key Words learning theory, conditioning, microeconomics, utility theory, uncertainty

■ **Abstract** The functions of rewards are based primarily on their effects on behavior and are less directly governed by the physics and chemistry of input events as in sensory systems. Therefore, the investigation of neural mechanisms underlying reward functions requires behavioral theories that can conceptualize the different effects of rewards on behavior. The scientific investigation of behavioral processes by animal learning theory and economic utility theory has produced a theoretical framework that can help to elucidate the neural correlates for reward functions in learning, goal-directed approach behavior, and decision making under uncertainty. Individual neurons can be studied in the reward systems of the brain, including dopamine neurons, orbitofrontal cortex, and striatum. The neural activity can be related to basic theoretical terms of reward and uncertainty, such as contiguity, contingency, prediction error, magnitude, probability, expected value, and variance.

CONTENTS

INTRODUCTION 88
GENERAL IDEAS ON REWARD FUNCTION, AND A CALL FOR THEORY 89
A Call for Behavioral Theory 91
REWARD FUNCTIONS DEFINED BY ANIMAL LEARNING THEORY 91
Learning 92
Approach Behavior 94
Motivational Valence 94
NEUROPHYSIOLOGY OF REWARD BASED ON ANIMAL LEARNING THEORY 95
Primary Reward 95
Contiguity 95
Contingency 95
Prediction Error 99
Approach Behavior and Goal Directedness 102
REWARD FUNCTIONS DEFINED BY MICROECONOMIC UTILITY THEORY 105

NEUROPHYSIOLOGY OF REWARD BASED ON ECONOMIC THEORY 107
Magnitude 107
Probability 108
Expected Value 109
Uncertainty 109
CONCLUSIONS 110

INTRODUCTION

How can we understand the common denominator of Pavlov's salivating dogs, an ale named Hobgoblin, a market in southern France, and the bargaining for lock access on the Mississippi River? Pavlov's dogs were presented with pieces of delicious sausage that undoubtedly made them salivate. We know that the same animal will salivate also when it hears a bell that has repeatedly sounded a few seconds before the sausage appears, as if the bell induced the well-known, pleasant anticipation of the desired sausage. Changing slightly the scenery, imagine you are in Cambridge, walk down Mill Lane, and unfailingly end up in the Mill pub by the river Cam. The known attraction inducing the pleasant anticipation is a pint of Hobgoblin. Hobgoblin's provocative ad reads something like "What's the matter Lager boy, afraid you might taste something?" and refers to a full-bodied, dark ale whose taste alone is a reward. Changing the scenery again, you are in the middle of a Saturday morning market in a small town in southern France and run into a nicely arranged stand of rosé and red wines. Knowing the presumably delicious contents of the differently priced bottles to varying degrees, you need to make a decision about what to get for lunch. You can do a numerical calculation and weigh the price of each bottle by the probability that its contents will please your taste, but chances are that a more automatic decision mechanism kicks in that is based on anticipation and will tell you quite quickly what to choose. However, you cannot use the same simple emotional judgment when you are in the shoes of an economist trying to optimize the access to the locks on the Mississippi River. The task is to find a pricing structure that assures the most efficient and uninterrupted use of the infrastructure over a 24-hour day, by avoiding long queues during prime daytime hours and inactive periods during the wee hours of the night. A proper pricing structure known in advance to the captains of the barges will shape their decisions to enter the locks at a moment that is economically most appropriate for the whole journey. The common denominator in these tasks appears to relate to the anticipation of outcomes of behavior in situations with varying degrees of uncertainty: the merely automatic salivation of a dog without much alternative, the choice of sophisticated but partly unknown liquids, or the well-calculated decision of a barge captain on how to get the most out of his money and time.

The performance in these tasks is managed by the brain, which assesses the values and uncertainties of predictable outcomes (sausage, ale, wine, lock pricing, and access to resources) and directs the individuals' decisions toward the current optimum. This review describes some of the knowledge on brain mechanisms related to rewarding outcomes, without attempting to provide a complete account of all the studies done. We focus on the activity of single neurons studied by neurophysiological techniques in behaving animals, in particular monkeys, and emphasize the formative role of behavioral theories, such as animal learning theory and microeconomic utility theory, on the understanding of these brain mechanisms. Given the space limits and the only just beginning neurophysiological studies based on game theory (Barraclough et al. 2004, Dorris & Glimcher 2004), the description of the neurophysiology of this promising field will have to wait until more data have been gathered. The review will not describe the neurobiology of artificial drug rewards, which constitutes a field of its own but does not require vastly different theoretical backgrounds of reward function for its understanding. Readers interested in the rapidly emerging and increasingly large field of human neuroimaging of reward and reward-directed decision making are referred to other reviews (O'Doherty 2004).

GENERAL IDEAS ON REWARD FUNCTION, AND A CALL FOR THEORY

Homer's Odysseus proclaims, "Whatever my distress may be, I would ask you now to let me eat. There is nothing more devoid of shame than the accursed belly; it thrusts itself upon a man's mind in spite of his afflictions. . .my heart is sad but my belly keeps urging me to have food and drink. . .it says imperiously: 'eat and be filled'." (*The Odyssey*, Book VII, 800 BC). Despite these suggestive words, Homer's description hardly fits the common-sensical perceptions of reward, which largely belong to one of two categories. People often consider a reward as a particular object or event that one receives for having done something well. You succeed in an endeavor, and you receive your reward. This reward function could be most easily accommodated within the framework of instrumental conditioning, according to which the reward serves as a positive reinforcer of a behavioral act. The second common perception of reward relates to subjective feelings of liking and pleasure. You do something again because it produced a pleasant outcome before. We refer to this as the hedonic function of rewards. The following descriptions will show that both of these perceptions of reward fall well short of providing a complete and coherent description of reward functions.

One of the earliest scientifically driven definitions of reward function comes from Pavlov (1927), who defined it as an object that produces a change in behavior, also called learning. The dog salivates to a bell only after the sound has been paired with a sausage, but not to a different, nonpaired sound, suggesting that its behavioral response (salivation) has changed after food conditioning. It is noteworthy that this definition bypasses both common-sensical reward notions, as the dog does not need to do anything in particular for the reward to occur (notion 1) nor is it relevant what the dog feels (notion 2). Yet we will see that this definition is a key to neurobiological studies.

Around this time, Thorndike's (1911) Law of Effect postulated that a reward increases the frequency and intensity of a specific behavioral act that has resulted in a reward before or, as a common interpretation has it, "rewards make you come back for more." This definition comes close to the idea of instrumental conditioning, in that you get a reward for having done something well, and not automatically as with Pavlovian conditioning. It resembles Pavlov's definition of learning function, as it suggests that you will do more of the same behavior that has led previously to the rewarding outcome (positive reinforcement). Skinner pushed the definition of instrumental, or operant, conditioning further by defining rewards as reinforcers of stimulus-response links that do not require mental processes such as intention, representation of goal, or consciousness. Although the explicit antimental stance reduced the impact of his concept, the purely behaviorist approach to studying reward function allowed scientists to acquire a huge body of knowledge by studying the behavior of animals, and it paved the way to neurobiological investigations without the confounds of subjective feelings.

Reward objects for animals are primarily vegetative in nature, such as different foodstuffs and liquids with various tastes. These rewards are necessary for survival, their motivational value can be determined by controlled access, and they can be delivered in quantifiable amounts in laboratory situations. The other main vegetative reward, sex, is impossible to deliver in neurophysiological laboratory situations requiring hundreds of daily trials. Animals are also sensitive to other, nonvegetative rewards, such as touch to the skin or fur and presentation of novel objects and situations eliciting exploratory responses, but these again are difficult to parameterize for laboratory situations. Humans use a wide range of nonvegetative rewards, such as money, challenge, acclaim, visual and acoustic beauty, power, security, and many others, but these are not considered as this review considers neural mechanisms in animals.

An issue with vegetative rewards is the precise definition of the rewarding effect. Is it the seeing of an apple, its taste on the tongue, the swallowing of a bite of it, the feeling of its going down the throat, or the rise in blood sugar subsequent to its digestion that makes it a reward and has one come back for more? Which of these events constitutes the primary rewarding effect, and do different objects draw their rewarding effects from different events (Wise 2002)? In some cases, the reward may be the taste experienced when an object activates the gustatory receptors, as with saccharin, which has no nutritional effects but increases behavioral reactions. The ultimate rewarding effect of many nutrient objects may be the specific influence on vegetative parameters, such as electrolyte, glucose, and amino acid concentrations in plasma and brain. This would explain why animals avoid foods that lack such nutrients as essential amino acids (Delaney & Gelperin 1986, Hrupka et al. 1997, Rogers & Harper 1970, Wang et al. 1996). The behavioral function of some reward objects may be determined by innate mechanisms, whereas a much larger variety might be learned through experience.

Although these theories provide important insights into reward function, they tend to neglect the fact that individuals usually operate in a world with limited nutritional and mating resources, and that most resources occur with different degrees of uncertainty. The animal in the wild is not certain whether it will encounter a particular fruit or prey object at a particular moment, nor is the restaurant goer certain that her preferred chef will cook that night. To make the uncertainty of outcomes tractable was the main motive that led Blaise Pascal to develop probability theory around 1650 (see Glimcher 2003 for details). He soon realized that humans make decisions by weighing the potential outcomes by their associated probabilities and then go for the largest result. Or, mathematically speaking, they sum the products of magnitude and probability of all potential outcomes of each option and then choose the option with the highest expected value. Nearly one hundred years later, Bernoulli (1738) discovered that the utility of outcomes for decision making does not increase linearly but frequently follows a concave function, which marks the beginning of microeconomic decision theory. The theory provides quantifiable assessments of outcomes under uncertainty and has gone a long way to explain human and animal decision making, even though more recent data cast doubt on the logic in some decision situations (Kahneman & Tversky 1984).

A Call for Behavioral Theory

Primary sensory systems have dedicated physical and chemical receptors that translate environmental energy and information into neural language. Thus, the functions of primary sensory systems are governed by the laws of mechanics, optics, acoustics, and receptor binding. By contrast, there are no dedicated receptors for reward, and the information enters the brain through mechanical, gustatory, visual, and auditory receptors of the sensory systems. The functions of rewards cannot be derived entirely from the physics and chemistry of input events but are based primarily on behavioral effects, and the investigation of reward functions requires behavioral theories that can conceptualize the different effects of rewards on behavior. Thus, the exploration of neural reward mechanisms should not be based primarily on the physics and chemistry of reward objects but on specific behavioral theories that define reward functions. Animal learning theory and microeconomics are two prominent examples of such behavioral theories and constitute the basis for this review.

REWARD FUNCTIONS DEFINED BY ANIMAL LEARNING THEORY

This section will combine some of the central tenets of animal learning theories in an attempt to define a coherent framework for the investigation of neural reward mechanisms. The framework is based on the description of observable behavior and superficially resembles the behaviorist approach, although mental states of representation and prediction are essential. Dropping the issues of subjective feelings of pleasure will allow us to do objective behavioral measurements in controlled neurophysiological experiments on animals. To induce subjective feelings of pleasure and positive emotion is a key function of rewards, although it is unclear whether the pleasure itself has a reinforcing, causal effect for behavior (i.e., I feel good because of the outcome I got and therefore will do again what produced the pleasant outcome) or is simply an epiphenomenon (i.e., my behavior gets reinforced and, in addition, I feel good because of the outcome).

Learning

Rewards induce changes in observable behavior and serve as positive reinforcers by increasing the frequency of the behavior that results in reward. In Pavlovian, or classical, conditioning, the outcome follows the conditioned stimulus (CS) irrespective of any behavioral reaction, and repeated pairing of stimuli with outcomes leads to a representation of the outcome that is evoked by the stimulus and elicits the behavioral reaction (Figure 1*a*). By contrast, instrumental, or operant, conditioning requires the subject to execute a behavioral response; without such response there will be no reward. Instrumental conditioning increases the frequency of those behaviors that are followed by reward by reinforcing stimulus-response links. Instrumental conditioning allows subjects to influence their environment and determine their rate of reward.

The behavioral reactions studied classically by Pavlov are vegetative responses governed by smooth muscle contraction and gland discharge, whereas more recent Pavlovian tasks also involve reactions of striated muscles. In the latter case, the final reward usually needs to be collected by an instrumental contraction of striated muscle, but the behavioral reaction to the CS itself, for example, anticipatory licking, is not required for the reward to occur and thus is classically conditioned. As a further emphasis on Pavlovian mechanisms, the individual stimuli in instrumental tasks that predict rewards are considered to be Pavlovian conditioned. These distinctions are helpful when trying to understand why the neural mechanisms of reward prediction reveal strong influences of Pavlovian conditioning.

Three factors govern conditioning, namely contiguity, contingency, and prediction error. Contiguity refers to the requirement of near simultaneity (Figure 1*a*). Specifically, a reward needs to follow a CS or response by an optimal interval of a few seconds, whereas rewards occurring before a stimulus or response do not contribute to learning (backward conditioning). The contingency requirement postulates that a reward needs to occur more frequently in the presence of a stimulus as compared with its absence in order to induce "excitatory" conditioning of the stimulus (Figure 1*b*); the occurrence of the CS predicts a higher incidence of reward compared with no stimulus, and the stimulus becomes a reward predictor. By contrast, if a reward occurs less frequently in the absence of a stimulus, compared with its presence, the occurrence of the stimulus predicts a lower incidence of reward, and the stimulus becomes a conditioned inhibitor, even though the contiguity

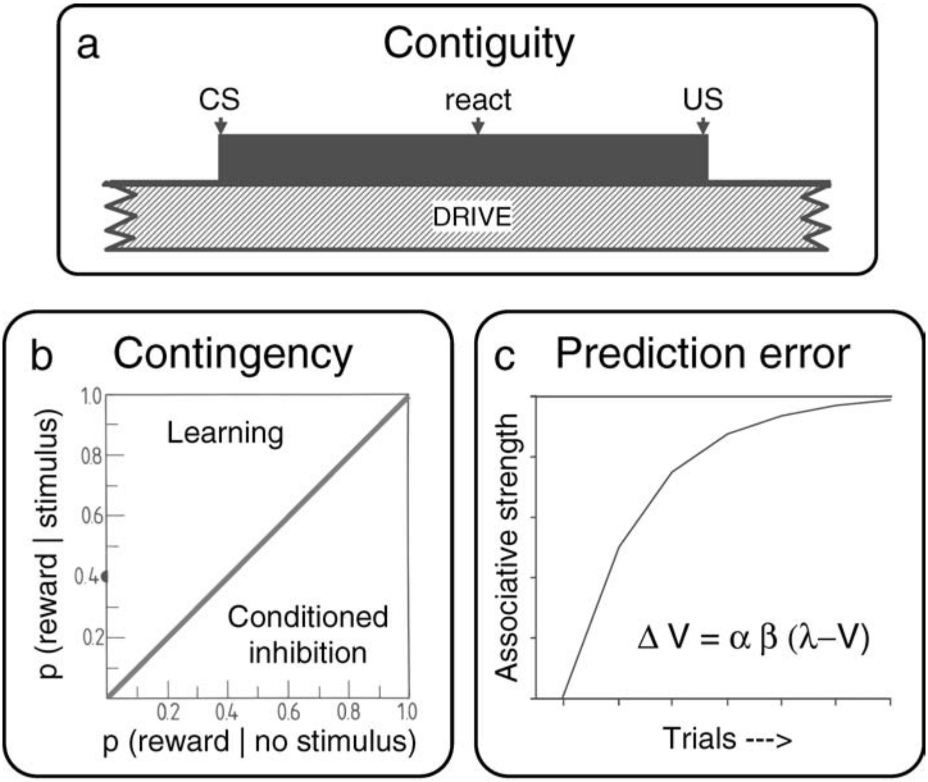


Figure 1 Basic assumptions of animal learning theory defining the behavioral functions of rewards. (*a*) Contiguity refers to the temporal proximity of a conditioned stimulus (CS), or action, and the reward. (*b*) Contingency refers to the conditional probability of reward occurring in the presence of a conditioned stimulus as opposed to its absence (modified from Dickinson 1980). (*c*) Prediction error denotes the discrepancy between an actually received reward and its prediction. Learning (ΔV, associative strength) is proportional to the prediction error (λ–V) and reaches its asymptote when the prediction error approaches zero after several learning trials. All three requirements need to be fulfilled for learning to occur. US, unconditioned stimulus.

requirement is fulfilled. The crucial role of prediction error is derived from Kamin's (1969) blocking effect, which postulates that a reward that is fully predicted does not contribute to learning, even when it occurs in a contiguous and contingent manner. This is conceptualized in the associative learning rules (Rescorla & Wagner 1972), according to which learning advances only to the extent to which a reinforcer is unpredicted and slows progressively as the reinforcer becomes more predicted (Figure 1*c*). The omission of a predicted reinforcer reduces the strength of the CS and produces extinction of behavior. So-called attentional learning rules in addition relate the capacity to learn (associability) in certain situations to the degree of attention evoked by the CS or reward (Mackintosh 1975, Pearce & Hall 1980).

Approach Behavior

Rewards elicit two forms of behavioral reactions, approach and consumption. This is because the objects are labeled with appetitive value through innate mechanisms (primary rewards) or, in most cases, classical or instrumental conditioning, after which these objects constitute, strictly speaking, conditioned reinforcers (Wise 2002). Nutritional rewards can derive their value from hunger and thirst states, and satiation of the animal reduces the reward value and consequently the behavioral reactions.

Conditioned, reward-predicting stimuli also induce preparatory or approach behavior toward the reward. In Pavlovian conditioning, subjects automatically show nonconsummatory behavioral reactions that would otherwise occur after the primary reward and that increase the chance of consuming the reward, as if a part of the behavioral response has been transferred from the primary reward to the CS (Pavlovian response transfer).

In instrumental conditioning, a reward can become a goal for instrumental behavior if two conditions are met. The goal needs to be represented at the time the behavior is being prepared and executed. This representation should contain a prediction of the future reward together with the contingency that associates the behavioral action to the reward (Dickinson & Balleine 1994). Behavioral tests for the role of "incentive" reward-predicting mechanisms include assessing behavioral performance in extinction following devaluation of the reward by satiation or aversive conditioning in the absence of the opportunity to perform the instrumental action (Balleine & Dickinson 1998). A reduction of behavior in this situation indicates that subjects have established an internal representation of the reward that is updated when the reward changes its value. (Performing the action together with the devalued outcome would result in reduced behavior due to partial extinction, as the reduced reward value would diminish the strength of the association.) To test the role of action-reward contingencies, the frequency of "free" rewards in the absence of the action can be varied to change the strength of association between the action and the reward and thereby modulate instrumental behavior (Balleine & Dickinson 1998).

Motivational Valence

Punishers have opposite valence to rewards, induce withdrawal behavior, and act as negative reinforcers by increasing the behavior that results in decreasing the aversive outcome. Avoidance can be passive when subjects increasingly refrain from doing something that is associated with a punisher (don't do it); active avoidance involves increasing an instrumental response that is likely to reduce the impact of a punisher (get away from it). Punishers induce negative emotional states of anger, fear, and panic.

NEUROPHYSIOLOGY OF REWARD BASED ON ANIMAL LEARNING THEORY

Primary Reward

Neurons responding to liquid or food rewards are found in a number of brain structures, such as orbitofrontal, premotor and prefrontal cortex, striatum, amygdala, and dopamine neurons (Amador et al. 2000, Apicella et al. 1991, Bowman et al. 1996, Hikosaka et al. 1989, Ljungberg et al. 1992, Markowitsch & Pritzel 1976, Nakamura et al. 1992, Nishijo et al. 1988, Pratt & Mizumori 1998, Ravel et al. 1999, Shidara et al. 1998, Thorpe et al. 1983, Tremblay & Schultz 1999). Satiation of the animal reduces the reward responses in orbitofrontal cortex (Critchley & Rolls 1996) and in the secondary gustatory area of caudal orbitofrontal cortex (Rolls et al. 1989), a finding that suggests that the responses reflect the rewarding functions of the objects and not their taste. Taste responses are found in the primary gustatory area of the insula and frontal operculum and are insensitive to satiation (Rolls et al. 1988).

Contiguity

Procedures involving Pavlovian conditioning provide simple paradigms for learning and allow the experimenter to test the basic requirements of contiguity, contingency, and prediction error. Contiguity can be tested by presenting a reward 1.5–2.0 seconds after an untrained, arbitrary visual or auditory stimulus for several trials. A dopamine neuron that responds initially to a liquid or food reward acquires a response to the CS after some tens of paired CS-reward trials (Figure 2) (Mirenowicz & Schultz 1994, Waelti 2000). Responses to conditioned, reward-predicting stimuli occur in all known reward structures of the brain, including the orbitofrontal cortex, striatum, and amygdala (e.g., Hassani et al. 2001, Liu & Richmond 2000, Nishijo et al. 1988, Rolls et al. 1996, Thorpe et al. 1983, Tremblay & Schultz 1999). (Figure 2 shows that the response to the reward itself disappears in dopamine neurons, but this is not a general phenomenon with other neurons.)

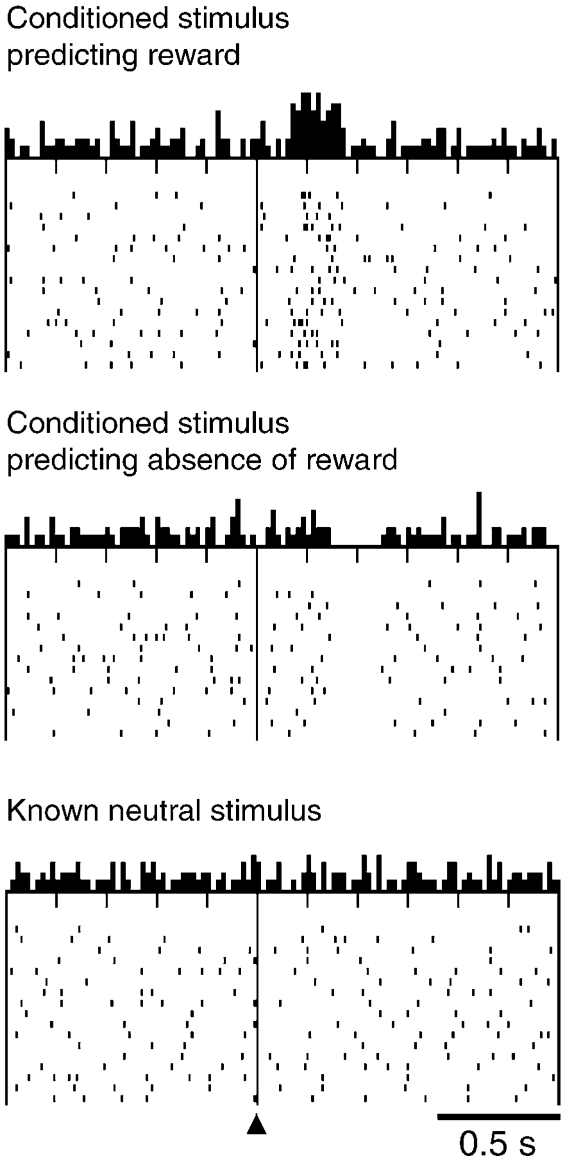
Contingency

The contingency requirement postulates that in order to be involved in reward prediction, neurons should discriminate between three kinds of stimuli, namely reward-predicting CSs (conditioned exciters), after which reward occurs more frequently compared with no CS (Figure 1*b,* top left); conditioned inhibitors, after which reward occurs less frequently compared with no CS (Figure 1*b*, bottom right); and neutral stimuli that are not associated with changes in reward frequency compared with no stimulus (diagonal line in Figure 1*b*). In agreement with these postulates, dopamine neurons are activated by reward-predicting CSs, show depressions of activity following conditioned inhibitors, which may be accompanied by small activations, and hardly respond to neutral stimuli when response generalization is excluded (Figure 3) (Tobler et al. 2003). The conditioned inhibitor in these experiments is set up by pairing the inhibitor with a reward-predicting CS while withholding the reward, which amounts to a lower probability of reward in the presence of the inhibitor compared with its absence (reward-predicting stimulus alone) and thus follows the scheme of Figure 1*b* (bottom right). Without conditioned inhibitors being tested, many studies find CS responses that distinguish between reward-predicting and neutral stimuli in all reward structures (e.g., Aosaki et al. 1994, Hollerman et al. 1998, Kawagoe et al. 1998, Kimura et al. 1984, Nishijo et al. 1988, Ravel et al. 1999, Shidara et al. 1998, Waelti et al. 2001).

Further tests assess the specificity of information contained in CS responses. In the typical behavioral tasks used in monkey experiments, the CS may contain several different stimulus components, namely spatial position; visual object features such as color, form, and spatial frequency; and motivational features such as reward prediction. It would be necessary to establish through behavioral testing which of these features is particularly effective in evoking a neural response. For example, neurons in the orbitofrontal cortex discriminate between different CSs on the basis of their prediction of different food and liquid rewards (Figure 4) (Critchley & Rolls 1996, Tremblay & Schultz 1999). By contrast, these neurons are less sensitive to the visual object features of the same CSs, and they rarely code their spatial position, although neurons in other parts of frontal cortex are particularly tuned to these nonreward parameters (Rao et al. 1997). CS responses that are primarily sensitive to the reward features are found also in the amygdala (Nishijo et al. 1988) and striatum (Hassani et al. 2001). These data suggest that individual neurons in these structures can extract the reward components from the multidimensional stimuli used in these experiments as well as in everyday life.

Reward neurons should distinguish rewards from punishers. Different neurons in orbitofrontal cortex respond to rewarding and aversive liquids (Thorpe et al. 1983). Dopamine neurons are activated preferentially by rewards and reward-predicting stimuli but are only rarely activated by aversive air puffs and saline (Mirenowicz & Schultz 1996). In anesthetized animals, dopamine neurons show depressions following painful stimuli (Schultz & Romo 1987, Ungless et al. 2004). Nucleus accumbens neurons in rats show differential activating or depressing responses to CSs predicting rewarding sucrose versus aversive quinine solutions in a Pavlovian task (Roitman et al. 2005). By contrast, the group of tonically active neurons of the striatum responds to both rewards and aversive air puffs, but not to neutral stimuli (Ravel et al. 1999). They seem to be sensitive to reinforcers in general, without specifying their valence. Alternatively, their responses might reflect the higher attention-inducing effects of reinforcers compared with neutral stimuli.

The omission of reward following a CS moves the contingency toward the diagonal line in Figure 1*b* and leads to extinction of learned behavior. By analogy, the withholding of reward reduces the activation of dopamine neurons by CSs within several tens of trials (Figure 5) (Tobler at al. 2003).

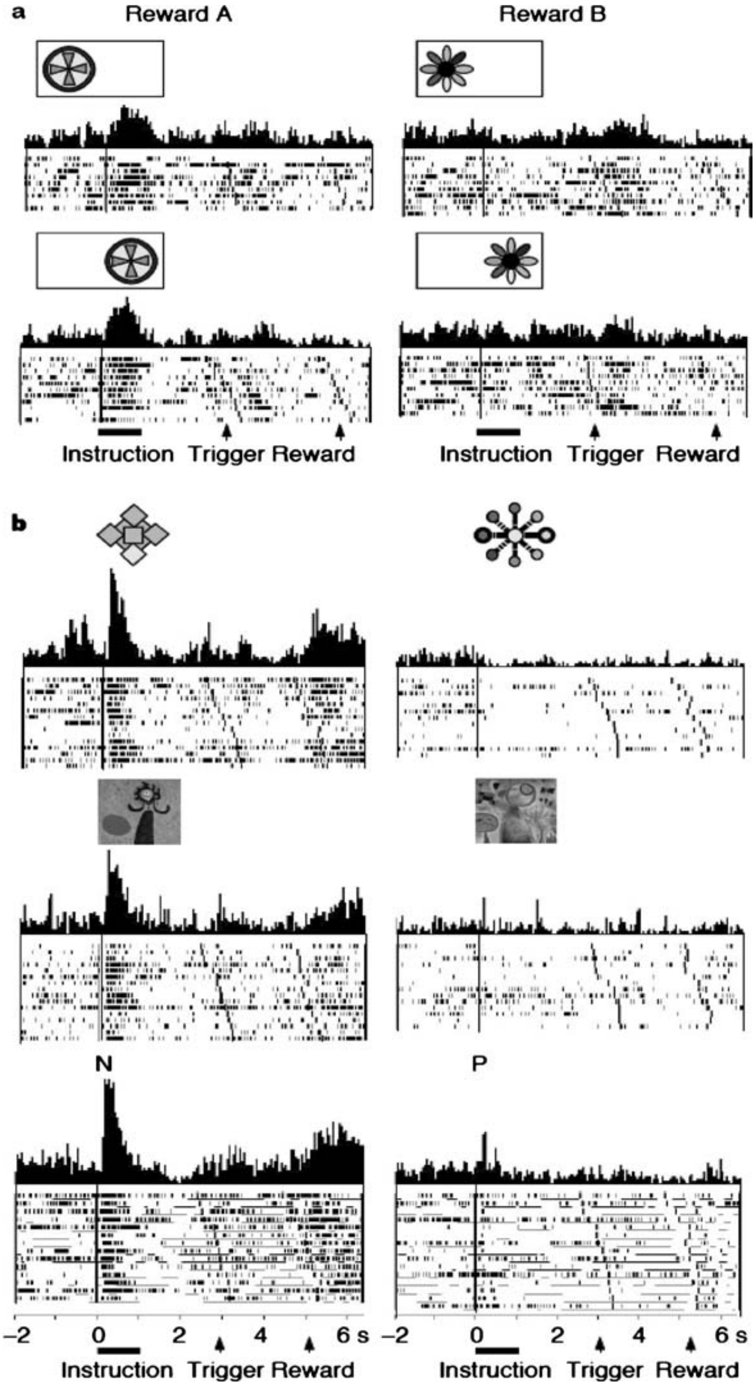


Conditioned stimulus predicting reward

Conditioned stimulus predicting absence of reward

Known neutral stimulus

Figure 3 Testing the contingency requirement for associative learning: responses of a single dopamine neuron to three types of stimuli. (*Top*) Activating response to a reward-predicting stimulus (higher occurrence of reward in the presence as opposed to absence of stimulus). (*Middle*) Depressant response to a different stimulus predicting the absence of reward (lower occurrence of reward in the presence as opposed to absence of stimulus). (*Bottom*) Neutral stimulus (no change in reward occurrence after stimulus). Vertical line and arrow indicate time of stimulus.


a
Reward A
Reward B
Instruction
Trigger
Reward
Instruction
Trigger
Reward
b
N
P
–2
0
2
4
6 s
–2
0
2
4
6 s
Instruction
Trigger
Reward
Instruction
Trigger
Reward

Prediction Error

Just as with behavioral learning, the acquisition of neuronal responses to reward-predicting CSs should depend on prediction errors. In the prediction error–defining blocking paradigm, dopamine neurons acquire a response to a CS only when the CS is associated with an unpredicted reward, but not when the CS is paired with a reward that is already predicted by another CS and the occurrence of the reward does not generate a prediction error (Figure 6) (Waelti et al. 2001). The neurons fail to learn to respond to reward predictors despite the fact that contiguity and contingency requirements for excitatory learning are fulfilled. These data demonstrate the crucial importance of prediction errors for associative neural learning and suggest that learning at the single-neuron level may follow similar rules as those for behavioral learning. This suggests that some behavioral learning functions may be carried by populations of single neurons.

Neurons may not only be sensitive to prediction errors during learning, but they may also emit a prediction error signal. Dopamine neurons, and some neurons in orbitofrontal cortex, show reward activations only when the reward occurs unpredictably and fail to respond to well-predicted rewards, and their activity is depressed when the predicted reward fails to occur (Figure 7) (Mirenowicz & Schultz 1994, Tremblay & Schultz 2000a). This result has prompted the notion that dopamine neurons emit a positive signal (activation) when an appetitive event is better than predicted, no signal (no change in activity) when an appetitive event occurs as predicted, and a negative signal (decreased activity) when an appetitive event is worse than predicted (Schultz et al. 1997). In contrast to this bidirectional error signal, some neurons in the prefrontal, anterior, and posterior cingulate cortex show a unidirectional error signal upon activation when a reward fails to occur because of a behavioral error of the animal (Ito et al. 2003, McCoy et al. 2003, Watanabe 1989; for review of neural prediction errors, see Schultz & Dickinson 2000).

More stringent tests for the neural coding of prediction errors include formal paradigms of animal learning theory in which prediction errors occur in specific situations. In the blocking paradigm, the blocked CS does not predict a reward. Accordingly, the absence of a reward following that stimulus does not produce a prediction error nor a response in dopamine neurons, and the delivery of a reward does produce a positive prediction error and a dopamine response (Figure 8*a*; left) (Waelti et al. 2001). By contrast, after a well-trained,

←

Figure 4 Reward discrimination in orbitofrontal cortex. (*a*) A neuron responding to the instruction cue predicting grenadine juice (*left*) but not apple juice (*right*), irrespective of the left or right position of the cue in front of the animal. (*b*) A different neuron responding to the cue predicting grape juice (*left*) but not orange juice (*right*), irrespective of the picture object predicting the juice. From Tremblay & Schultz 1999, © Nature MacMillan Publishers.

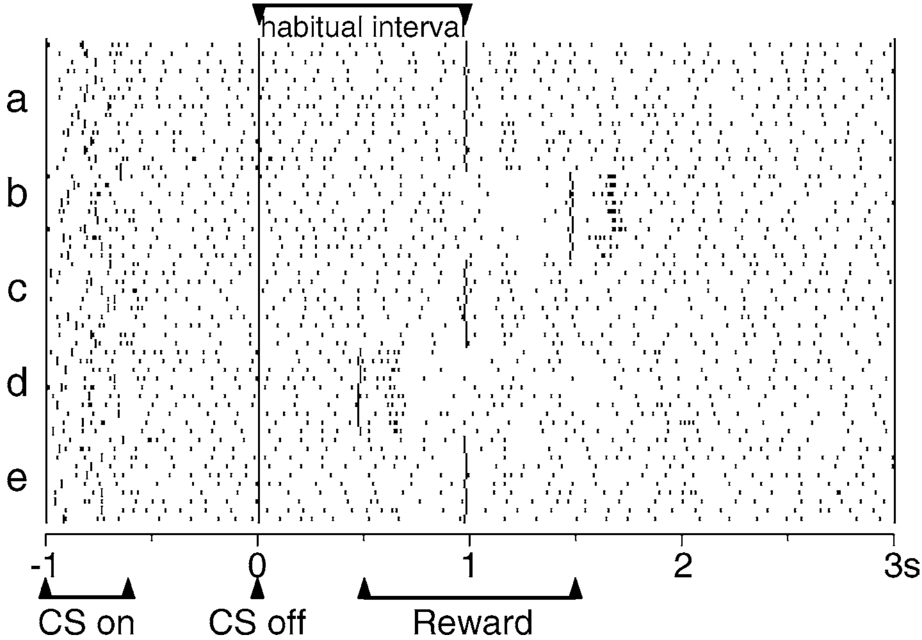


Figure 7 Dopamine response codes temporal reward prediction error. (*a*, *c*, *e*) No response to reward delivered at habitual time. (*b*) Delay in reward induces depression at previous time of reward, and activation at new reward time. (*d*) Precocious reward delivery induces activation at new reward time, but no depression at previous reward time. Trial sequence is from top to bottom. Data from Hollerman & Schultz (1998). CS, conditioned stimulus.

reward-predicting CS, reward omission produces a negative prediction error and a depressant neural response, and reward delivery does not lead to a prediction error or a response in the same dopamine neuron (Figure 8*a*; right). In a conditioned inhibition paradigm, the conditioned inhibitor predicts the absence of reward, and the absence of reward after this stimulus does not produce a prediction error or a response in dopamine neurons, even when another, otherwise reward-predicting stimulus is added (Figure 8*b*) (Tobler at al. 2003). By contrast, the occurrence of reward after an inhibitor produces an enhanced prediction error, as the prediction error represents the difference between the actual reward and the negative prediction from the inhibitor, and the dopamine neuron shows a strong response (Figure 8*b*; bottom). Taken together, these data suggest that dopamine neurons show bidirectional coding of reward prediction errors, following the equation

$$\text{Dopamine response} = \text{Reward occurred} - \text{Reward predicted}.$$

This equation may constitute a neural equivalent for the prediction error term of (λ–V) of the Rescorla-Wagner learning rule. With these characteristics, the bidirectional dopamine error response would constitute an ideal teaching signal for neural plasticity.

The neural prediction error signal provides an additional means to investigate the kinds of information contained in the representations evoked by CSs. Time apparently plays a major role in behavioral learning, as demonstrated by the unblocking effects of temporal variations of reinforcement (Dickinson et al. 1976). Figure 7 shows that the prediction acting on dopamine neurons concerns the exact time of reward occurrence. Temporal deviations induce a depression when the reward fails to occur at the predicted time (time-sensitive reward omission response), and an activation when the reward occurs at a moment other than predicted (Hollerman & Schultz 1998). This time sensitivity also explains why neural prediction errors occur at all in the laboratory in which animals know that they will receive ample quantities of reward but without knowing when exactly the reward will occur. Another form of time representation is revealed by tests in which the probability of receiving a reward after the last reward increases over consecutive trials. Thus, the animal's reward prediction should increase after each unrewarded trial, the positive prediction error with reward should decrease, and the negative prediction error with reward omission should increase. In line with this reasoning, dopamine neurons show progressively decreasing activations to reward delivery as the number of trials since the last reward increases, and increasing depressions in unrewarded trials (Figure 9) (Nakahara et al. 2004). The result suggests that, for the neurons, the reward prediction in the CS increases after every unrewarded trial, due to the temporal profile of the task evoked by the CS, and contradicts an assumption from temporal difference reinforcement modeling that the prediction error of the preceding unrewarded trial would reduce the current reward prediction in the CS, in which case the neural prediction error responses should increase, which is the opposite to what is actually observed (although the authors attribute the temporal conditioning to the context and have the CS conform to the temporal difference model). The results from the two experiments demonstrate that dopamine neurons are sensitive to different aspects of temporal information evoked by reward-predicting CSs and demonstrate how experiments based on specific behavioral concepts, namely prediction error, reveal important characteristics of neural coding.

The uncertainty of reward is a major factor for generating the attention that determines learning according to the associability learning rules (Mackintosh 1975, Pearce & Hall 1980). When varying the probability of reward in individual trials from 0 to 1, reward becomes most uncertain at $p = 0.5$, as it is most unclear whether or not a reward will occur. (Common perception might say that reward is even more uncertain at $p = 0.25$; however, at this low probability, it is nearly certain that reward will not occur.) Dopamine neurons show a slowly increasing activation between the CS and reward that is maximal at $p = 0.5$ (Fiorillo et al. 2003). This response may constitute an explicit uncertainty signal and is different in time and occurrence from the prediction error response. The response might contribute to a teaching signal in situations defined by the associability learning rules.

Approach Behavior and Goal Directedness

Many behavioral tasks in the laboratory involve more than a CS and a reward and comprise instrumental ocular or skeletal reactions, mnemonic delays between instruction cues and behavioral reactions, and delays between behavioral reactions and rewards during which animals can expect the reward.

Appropriately conditioned stimuli can evoke specific expectations of reward, and phasic neural responses to these CSs may reflect the process of evocation (see above). Once the representations have been evoked, their content can influence the behavior during some time. Neurons in a number of brain structures show sustained activations after an initial CS has occurred. The activations arise usually during specific epochs of well-differentiated instrumental tasks, such as during movement preparation (Figure 10*a*) and immediately preceding the reward (Figure 10*b*), whereas few activations last during the entire period between CS and reward. The activations differentiate between reward and no reward, between different kinds of liquid and food reward, and between different magnitudes of reward. They occur in all trial types in which reward is expected, irrespective of the type of behavioral action (Figure 10). Thus, the activations appear to represent reward expectations. They are found in the striatum (caudate, putamen, ventral

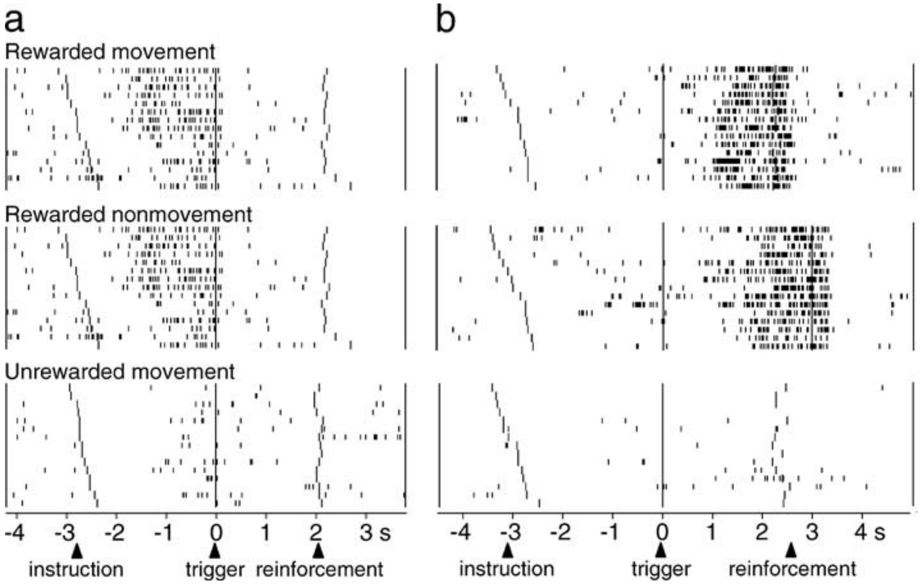


Figure 10 Reward expectation in the striatum. (*a*) Activation in a caudate neuron preceding the stimulus that triggers the movement or nonmovement reaction in both rewarded trial types irrespective of movement, but not in unrewarded movement trials. (*b*) Activation in a putamen neuron preceding the delivery of liquid reward in both rewarded trial types, but not before the reinforcing sound in unrewarded movement trials. Data from Hollerman et al. (1998).

striatum), amygdala, orbitofrontal cortex, dorsolateral prefrontal cortex, anterior cingulate, and supplementary eye field (Amador et al. 2000, Apicella et al. 1992, Cromwell & Schultz 2003, Hikosaka et al. 1989, Hollerman et al. 1998, Pratt & Mizumori 2001, Schoenbaum et al. 1998, Schultz et al. 1992, Shidara & Richmond 2002, Tremblay & Schultz 1999, 2000a, Watanabe 1996, Watanabe et al. 2002). Reward expectation-related activity in orbitofrontal cortex and amygdala develops as the reward becomes predictable during learning (Schoenbaum et al. 1999). In learning episodes with pre-existing reward expectations, orbitofrontal and striatal activations occur initially in all situations but adapt to the currently valid expectations, for example when novel stimuli come to indicate rewarded versus unrewarded trials. The neural changes occur in parallel with the animal's behavioral differentiation (Tremblay et al. 1998, Tremblay & Schultz 2000b).

In some neurons, the differential reward expectation-related activity discriminates in addition between different behavioral responses, such as eye and limb movements toward different spatial targets and movement versus nonmovement reactions (Figure 11). Such neurons are found in the dorsolateral prefrontal cortex (Kobayashi et al. 2002, Matsumoto et al. 2003, Watanabe 1996) and striatum

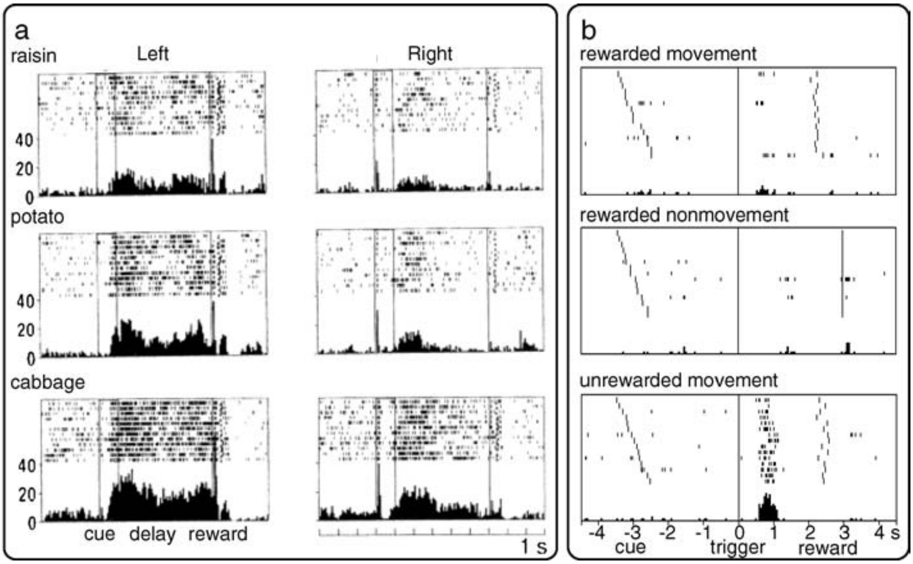


Figure 11 Potential neural mechanisms underlying goal-directed behavior. (*a*) Delay activity of a neuron in primate prefrontal cortex that encodes, while the movement is being prepared, both the behavioral reaction (left versus right targets) and the kind of outcome obtained for performing the action. From Watanabe (1996), © Nature MacMillan Publishers. (*b*) Response of a caudate neuron to the movement-triggering stimulus exclusively in unrewarded trials, thus coding both the behavioral reaction being executed and the anticipated outcome of the reaction. Data from Hollerman et al. (1998).

(Cromwell & Schultz 2003, Hassani et al. 2001, Hollerman et al. 1998, Kawagoe et al. 1998). The activations occur during task epochs related to the preparation and execution of the movement that is performed in order to obtain the reward. They do not simply represent outcome expectation, as they differentiate between different behavioral reactions despite the same outcome (Figure 11*a*, left versus right; Figure 11*b*, movement versus nonmovement), and they do not simply reflect different behavioral reactions, as they differentiate between the expected outcomes (Figure 11*a*,*b*, top versus bottom). Or, expressed in another way, the neurons show differential, behavior-related activations that depend on the outcome of the trial, namely reward or no reward and different kinds and magnitudes of reward. The differential nature of the activations develops during learning while the different reward expectations are being acquired, similar to simple reward expectation-related activity (Tremblay et al. 1998).

It is known that rewards have strong attention-inducing functions, and reward-related activity in parietal association cortex might simply reflect the known involvement of these areas in attention (Maunsell 2004). It is often tedious to disentangle attention from reward, but one viable solution would be to test neurons for specificity for reinforcers with opposing valence while keeping the levels of reinforcement strength similar for rewards and punishers. The results of such tests suggest that dopamine neurons and some neurons in orbitofrontal cortex discriminate between rewards and aversive events and thus report reward-related but not attention-related stimulus components (Mirenowicz & Schultz 1996, Thorpe et al. 1983). Also, neurons showing increasing activations with decreasing reward value or magnitude are unlikely to reflect the attention associated with stronger rewards. Such inversely related neurons exist in the striatum and orbitofrontal cortex (Hassani et al. 2001, Hollerman et al. 1998, Kawagoe et al. 1998, Watanabe 1996).

General learning theory suggests that Pavlovian associations of reward-predicting stimuli in instrumental tasks relate either to explicit CSs or to contexts. The neural correlates of behavioral associations with explicit stimuli may not only involve the phasic responses to CSs described above but also activations at other task epochs. Further neural correlates of Pavlovian conditioning may consist of the sustained activations that occur during the different task periods preceding movements or rewards (Figure 10), which are only sensitive to reward parameters and not to the types of behavioral reactions necessary to obtain the rewards.

Theories of goal-directed instrumental behavior postulate that in order to consider rewards as goals of behavior, there should be (*a*) an expectation of the outcome at the time of the behavior that leads to the reward, and (*b*) a representation of the contingency between the instrumental action and the outcome (Dickinson & Balleine 1994). The sustained, reward-discriminating activations may constitute a neural mechanism for simple reward expectation, as they reflect the expected reward without differentiating between behavioral reactions (Figure 10). However, these activations are not fully sufficient correlates for goal-directed behavior, as the reward expectation is not necessarily related to the specific action that results in the goal being attained; rather, it might refer to an unrelated reward that occurs in parallel and irrespective of the action. Such a reward would not constitute a goal of the action, and the reward-expecting activation might simply reflect the upcoming reward without being involved in any goal mechanism. By contrast, reward-expecting activations might fulfill the second, more stringent criterion if they are also specific for the action necessary to obtain the reward. These reward-expecting activations differentiate between different behavioral acts and arise only under the condition that the behavior leading to the reward is being prepared or executed (Figure 11). Mechanistically speaking, the observed neural activations may be the result of convergent neural coding of reward and behavior, but from a theoretical point, the activations could represent evidence for neural correlates of goal-directed mechanisms. To distinguish between the two possibilities, it would be helpful to test explicitly the contingency requirement by varying the probabilities of reward in the presence versus absence of behavioral reactions. Further tests could employ reward devaluations to distinguish between goal-directed and habit mechanisms, as the relatively more simple habits might also rely on combined neural mechanisms of expected reward and behavioral action but lack the more flexible representations of reward that are the hallmark of goal mechanisms.

REWARD FUNCTIONS DEFINED BY MICROECONOMIC UTILITY THEORY

How can we compare apples and pears? We need a numerical scale in order to assess the influence of different rewards on behavior. A good way to quantify the value of individual rewards is to compare them in choice behavior. Given two options, I would choose the one that at this moment has the higher value for me. Give me the choice between a one-dollar bill and an apple, and you will see which one I prefer and thus my action will tell you whether the value of the apple for me is higher or lower or similar compared with one dollar. To be able to put a quantifiable, numerical value onto every reward, even when the value is short-lived, has enormous advantages for getting reward-related behavior under experimental control.

To obtain a more complete picture, we need to take into account the uncertainty with which rewards frequently occur. One possibility would be to weigh the value of individual rewards with the probability with which they occur, an approach taken by Pascal ca. 1650. The sum of the products of each potential reward and its probability defines the expected value (EV) of the probability distribution and thus the theoretically expected payoff of an option, according to

$$\mathrm{EV} = \sum_{\mathrm{i}} (\mathrm{p_i} \cdot \mathrm{x_i});\ \ \mathrm{i} = 1, \mathrm{n};\ \mathrm{n} = \text{number of rewards}.$$

With increasing numbers of trials, the measured mean of the actually occurring distribution will approach the expected value. Pascal conjectured that human choice behavior could be approximated by this procedure.

Despite its advantages, expected value theory has limits when comparing very small with very large rewards or when comparing values at different start positions. Rather than following physical sizes of reward value in a linear fashion, human choice behavior in many instances increases more slowly as the values get higher, and the term of utility, or in some cases prospect, replaces the term of value when the impact of rewards on choices is assessed (Bernoulli 1738, Kahneman & Tversky 1984, Savage 1954, von Neumann & Morgenstern 1944). The utility function can be modeled by various equations (for detailed descriptions, see Gintis 2000, Huang & Litzenberger 1988), such as

1. The logarithmic utility function, $u(x) = \ln(x)$, yields a concave curve similar to the Weber (1850) function of psychophysics.
2. The power utility function, $u(x) = x^a$. With a € (0,1), and often a € [0.66, 0.75], the function is concave and resembles the power law of psychophysics (Stevens 1957). By contrast, $a = 1.0$ produces a linear function in which utility (value) = value. With $a > 1$, the curve becomes convex and increases faster toward higher values.
3. The exponential utility function, $u(x) = 1 - e^{-b\,x}$, produces a concave function for b € (0,1).
4. With the weighted reward value being expressed as utility, the expected value of a gamble becomes the expected utility (EU) according to

$$\mathrm{EU} = \sum_i (p_i \cdot u(x_i));\ i = 1, n;\ n = \text{number of rewards.}$$

Assessing the expected utility allows comparisons between gambles that have several outcomes with different values occurring at different probabilities. Note that a gamble with a single reward occurring at a $p < 1$ actually has two outcomes, the reward occurring with p and the nonreward with $(1 - p)$. A gamble with only one reward at $p = 1.0$ is called a safe option. Risk refers simply to known probabilities of < 1.0 and does not necessarily involve loss. Risky gambles have known probabilities; ambiguous gambles have probabilities unknown to the agent.

The shape of the utility function allows us to deal with the influence of uncertainty on decision-making. Let us assume an agent whose decision making is characterized by a concave utility function, as shown in Figure 12, who performs in a gamble with two outcomes of values 1 and 9 at $p = 0.5$ each (either the lower or the higher outcome will occur, with equal probability). The EV of the gamble is 5 (vertical dotted line), and the utility u(EV) (horizontal dotted line) lies between u(1) and u(9) (horizontal lines). Interestingly, u(EV) lies closer to u(9) than to u(1), suggesting that the agent foregoes more utility when the gamble produces u(1) than she wins with u(9) over u(EV). Given that outcomes 1 and 9 occur with the same frequency, this agent would profit more from a safe reward at EV, with u(EV), over the gamble. She should be risk averse. Thus, a concave utility function suggests risk aversion, whereas a convex function, in which an agent foregoes less reward than she wins, suggests risk seeking. Different agents with different attitudes toward risk have differently shaped utility functions.

A direct measure of the influence of uncertainty is obtained by considering the difference between u(EV) and the EU of the gamble. The EU in the case of equal probabilities is the mean of u(1) and u(9), as marked by EU(1–9), which is considerably lower than u(EV) and thus indicates the loss in utility due to risk. By comparison, the gamble of 4 and 6 involves a smaller range of reward magnitudes and thus less risk and less loss due to uncertainty, as seen by comparing the vertical bars associated with EU(4–6) and EU(1–9). This graphical analysis suggests that value and uncertainty of outcome can be considered as separable measures.

A separation of value and uncertainty as components of utility can be achieved mathematically by using, for example, the negative exponential utility function often employed in financial mathematics. Using the exponential utility function for EU results in

$$\mathrm{EU} = \sum_{\mathrm{i}} (\mathrm{p_i} \cdot -\mathrm{e}^{-\mathrm{b\,x}}\mathrm{i}),$$

which can be developed by the Laplace transform into

$$\mathrm{EU} = -\mathrm{e}^{-\mathrm{b\,(EV - b/2 \cdot var)}},$$

where EV is expected value, var is variance, and the probability distribution pi is Gaussian. Thus, EU is expressed as f(EV, variance). This procedure uses variance as a measure of uncertainty. Another measure of uncertainty is the entropy of information theory, which might be appropriate to use when dealing with information processing in neural systems, but entropy is not commonly employed for describing decision making in microeconomics.

Taken together, microeconomic utility theory has defined basic reward parameters, such as magnitude, probability, expected value, expected utility, and variance, that can be used for neurobiological experiments searching for neural correlates of decision making under uncertainty.

NEUROPHYSIOLOGY OF REWARD BASED ON ECONOMIC THEORY

Magnitude

The easiest quantifiable measure of reward for animals is the volume of juice, which animals can discriminate in submilliliter quantities (Tobler et al. 2005). Neurons show increasing responses to reward-predicting CSs with higher volumes of reward in a number of reward structures, such as the striatum (Cromwell & Schultz 2003) (Figure 13*a*), dorsolateral and orbital prefrontal cortex (Leon & Shadlen 1999, Roesch & Olson 2004, Wallis & Miller 2003), parietal and posterior cingulate cortex (McCoy et al. 2003, Musallam et al. 2004, Platt & Glimcher 1999), and dopamine neurons (Satoh et al. 2003, Tobler et al. 2005). Similar reward magnitude–discriminating activations are found in these structures in relation to other task events, before and after reward delivery. Many of these studies also report decreasing responses with increasing reward magnitude (Figure 13*b*), although not with dopamine neurons. The decreasing responses are likely to reflect true magnitude discrimination rather than simply the attention induced by rewards, which should increase with increasing magnitude.

Recent considerations cast doubt on the nature of some of the reward magnitude–discriminating, behavior-related activations, in particular in structures involved in motor and attentional processes, such as the premotor cortex, frontal eye fields, supplementary eye fields, parietal association cortex, and striatum. Some reward-related differences in movement-related activations might reflect the differences in movements elicited by different reward magnitudes (Lauwereyns et al. 2002, Roesch & Olson 2004). A larger reward might make the animal move faster, and increased neural activity in premotor cortex with larger reward might reflect the higher movement speed. Although a useful explanation for motor structures, the issue might be more difficult to resolve for areas more remote from motor output, such as prefrontal cortex, parietal cortex, and caudate nucleus. It would be helpful to correlate reward magnitude–discriminating activity in single neurons with movement parameters, such as reaction time and movement speed, and, separately, with reward parameters, and see where higher correlations are obtained. However, the usually measured movement parameters may not be sensitive enough to make these distinctions when neural activity varies relatively little with reward magnitude. On the other hand, inverse relationships, such as higher neural activity for slower movements associated with smaller rewards, would argue against a primarily motor origin of reward-related differences, as relatively few neurons show higher activity with slower movements.

Probability

Simple tests for reward probability involve CSs that differentially predict the probability with which a reward, as opposed to no reward, will be delivered for trial completion in Pavlovian or instrumental tasks. Dopamine neurons show increasing phasic responses to CSs that predict reward with increasing probability (Fiorillo et al. 2003, Morris et al. 2004). Similar increases in task-related activity occur in parietal cortex and globus pallidus during memory and movement-related task periods (Arkadir et al. 2004, Musallam et al. 2004, Platt & Glimcher 1999). However, reward-responsive tonically active neurons in the striatum do not appear to be sensitive to reward probability (Morris et al. 2004), indicating that not all neurons sensitive to reward may code its value in terms of probability. In a decision-making situation with varying reward probabilities, parietal neurons track the recently experienced reward value, indicating a memory process that would provide important input information for decision making (Sugrue et al. 2004).

Expected Value

Parietal neurons show increasing task-related activations with both the magnitude and probability of reward that do not seem to distinguish between the two components of expected value (Musallam et al. 2004). When the two value parameters are tested separately and in combination, dopamine neurons show monotonically increasing responses to CSs that predict increasing value (Tobler et al. 2005). The neurons fail to distinguish between magnitude and probability and seem to code their product (Figure 14*a*). However, the neural noise inherent in the stimulus-response relationships makes it difficult to determine exactly whether dopamine neurons encode expected value or expected utility. In either case, it appears as if neural responses show a good relationship to theoretical notions of outcome value that form a basis for decision making.

Uncertainty

Graphical analysis and application of the Laplace transform on the exponential utility function would permit experimenters to separate the components of expected value and utility from the uncertainty inherent in probabilistic gambles. Would the

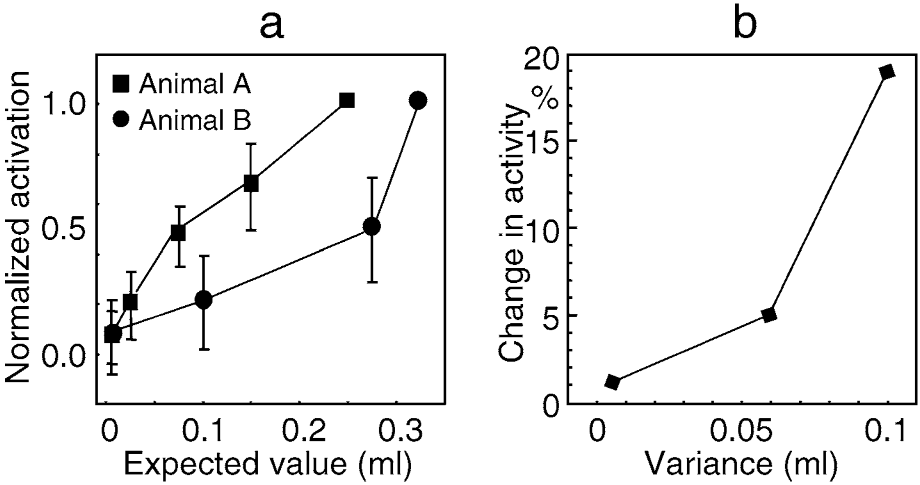


Figure 14 Separate coding of reward value and uncertainty in dopamine neurons. (*a*) Phasic response to conditioned, reward-predicting stimuli scales with increasing expected value (EV, summed magnitude × probability). Data points represent median responses normalized to response to highest EV (animal A, 57 neurons; animal B, 53 neurons). Data from Tobler et al. (2005). (*b*) Sustained activation during conditioned stimulus–reward interval scales with increasing uncertainty, as measured by variance. Two reward magnitudes are delivered at p = 0.5 each (0.05–0.15, 0.15–0.5 ml, 0.05–0.5 ml). Ordinate shows medians of changes above background activity from 53 neurons. Note that the entropy stays 1 bit for all three probability distributions. Data from Fiorillo et al. (2003).

brain be able to produce an explicit signal that reflects the level of uncertainty, similar to producing a reward signal? For both reward and uncertainty, there are no specialized sensory receptors. A proportion of dopamine neurons show a sustained activation during the CS-reward interval when tested with CSs that predict reward at increasing probabilities, as opposed to no reward. The activation is highest for reward at $p = 0.5$ and progressively lower for probabilities further away from $p = 0.5$ in either direction (Fiorillo et al. 2003). The activation does not occur when reward is substituted by a visual stimulus. The activations appear to follow common measures of uncertainty, such as statistical variance and entropy, both of which are maximal at $p = 0.5$. Most of the dopamine neurons signaling reward uncertainty also show phasic responses to reward-predicting CSs that encode expected value, and the two responses coding different reward terms are not correlated with each other. When in a refined experiment two different reward magnitudes alternate randomly (each at $p = 0.5$), dopamine neurons show the highest sustained activation when the reward range is largest, indicating a relationship to the statistical variance and thus to the uncertainty of the reward (Figure 14*b*). In a somewhat comparable experiment, neurons in posterior cingulate cortex show increased task-related activations as animals choose among rewards with larger variance compared with safe options (McCoy & Platt 2003). Although only a beginning, these data suggest that indeed the brain may produce an uncertainty signal about rewards that could provide essential information when making decisions under uncertainty. The data on dopamine neurons suggest that the brain may code the expected value separately from the uncertainty, just as the two terms constitute separable components of expected utility when applying the Laplace transform on the exponential utility function.

CONCLUSIONS

It is intuitively simple to understand that the use of well-established behavioral theories can only be beneficial when working with mechanisms underlying behavioral reactions. Indeed, these theories can very well define the different functions of rewards on behavior. It is then a small step on firm ground to base the investigation of neural mechanisms underlying the different reward functions onto the phenomena characterized by these theories. Although each theory has its own particular emphasis, they deal with the same kinds of outcome events of behavior, and it is more confirmation than surprise to see that many neural reward mechanisms can be commonly based on, and understood with, several theories. For the experimenter, the use of different theories provides good explanations for an interesting spectrum of reward functions that may not be so easily accessible by using only a single theory. For example, it seems that uncertainty plays a larger role in parts of microeconomic theory than in learning theory, and the investigation of neural mechanisms of uncertainty in outcomes of behavior can rely on several hundred years of thoughts about decision making (Pascal 1650 in Glimcher 2003, Bernoulli 1738).

ACKNOWLEDGMENTS

This article is based on lectures delivered at a Society for Neuroscience Meeting in October 2004 in San Diego and on a Max-Planck Symposium in January 2005 in Frankfurt, Germany. The author wishes to thank Drs. Anthony Dickinson and Peter Bossaerts for illuminating discussions on behavioral theories. Our work was supported by grants from the Wellcome Trust, Swiss National Science Foundation, Human Frontiers, and European Community.

The *Annual Review of Psychology* is online at http://psych.annualreviews.org

LITERATURE CITED

Amador N, Schlag-Rey M, Schlag J. 2000. Reward-predicting and reward-detecting neuronal activity in the primate supplementary eye field. *J. Neurophysiol.* 84:2166–70

Aosaki T, Tsubokawa H, Ishida A, Watanabe K, Graybiel AM, Kimura M. 1994. Responses of tonically active neurons in the primate's striatum undergo systematic changes during behavioral sensorimotor conditioning. *J. Neurosci.* 14:3969–84

Apicella P, Ljungberg T, Scarnati E, Schultz W. 1991. Responses to reward in monkey dorsal and ventral striatum. *Exp. Brain Res.* 85:491–500

Apicella P, Scarnati E, Ljungberg T, Schultz W. 1992. Neuronal activity in monkey striatum related to the expectation of predictable environmental events. *J. Neurophysiol.* 68:945–60

Arkadir D, Morris G, Vaadia E, Bergman H. 2004. Independent coding of movement direction and reward prediction by single pallidal neurons. *J. Neurosci.* 24:10047–56

Balleine B, Dickinson A. 1998. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* 37:407–19

Barraclough D, Conroy ML, Lee DJ. 2004. Prefrontal cortex and decision making in a mixed-strategy game. *Nat. Neurosci.* 7:405–10

Bernoulli J. (1738) 1954. Exposition of a new theory on the measurement of risk. *Econometrica* 22:23–36

Bowman EM, Aigner TG, Richmond BJ. 1996. Neural signals in the monkey ventral striatum related to motivation for juice and cocaine rewards. *J. Neurophysiol.* 75:1061–73

Critchley HG, Rolls ET. 1996. Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex. *J. Neurophysiol.* 75:1673–86

Cromwell HC, Schultz W. 2003. Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *J. Neurophysiol.* 89:2823–38

Delaney K, Gelperin A. 1986. Post-ingestive food-aversion learning to amino acid deficient diets by the terrestrial slug *Limax maximus*. *J. Comp. Physiol. A* 159:281–95

Dickinson A. 1980. *Contemporary Animal Learning Theory*. Cambridge, UK: Cambridge Univ. Press

Dickinson A, Balleine B. 1994. Motivational control of goal-directed action. *Anim. Learn. Behav.* 22:1–18

Dickinson A, Hall G, Mackintosh NJ. 1976. Surprise and the attenuation of blocking. *J. Exp. Psychol. Anim. Behav. Process.* 2:313–22

Dorris MC, Glimcher PW. 2004. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* 44:365–78

Fiorillo CD, Tobler PN, Schultz W. 2003. Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299:1898–902

Gintis H. 2000. *Game Theory Evolving*. Princeton, NJ: Princeton Univ. Press

Glimcher PW. 2003. *Decisions, Uncertainty and the Brain*. Cambridge, MA: MIT Press

Hassani OK, Cromwell HC, Schultz W. 2001. Influence of expectation of different rewards on behavior-related neuronal activity in the striatum. *J. Neurophysiol.* 85:2477–89

Hikosaka K, Watanabe M. 2000. Delay activity of orbital and lateral prefrontal neurons of the monkey varying with different rewards. *Cereb. Cortex* 10:263–71

Hikosaka O, Sakamoto M, Usui S. 1989. Functional properties of monkey caudate neurons. III. Activities related to expectation of target and reward. *J. Neurophysiol.* 61:814–32

Hollerman JR, Schultz W. 1998. Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* 1:304–9

Hollerman JR, Tremblay L, Schultz W. 1998. Influence of reward expectation on behavior-related neuronal activity in primate striatum. *J. Neurophysiol.* 80:947–63

Hrupka BJ, Lin YM, Gietzen DW, Rogers QR. 1997. Small changes in essential amino acid concentrations alter diet selection in amino acid-deficient rats. *J. Nutr.* 127:777–84

Huang C-F, Litzenberger RH. 1988. *Foundations for Financial Economics*. Upper Saddle River, NJ: Prentice Hall

Ito S, Stuphorn V, Brown JW, Schall JD. 2003. Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* 302:120–22

Kahneman D, Tversky A. 1984. Choices, values, and frames. *Am. Psychol.* 4:341–50

Kamin LJ. 1969. Selective association and conditioning. In *Fundamental Issues in Instrumental Learning*, ed. NJ Mackintosh, WK Honig, pp. 42–64. Halifax, NS: Dalhousie Univ. Press

Kawagoe R, Takikawa Y, Hikosaka O. 1998. Expectation of reward modulates cognitive signals in the basal ganglia. *Nat. Neurosci.* 1:411–16

Kimura M, Rajkowski J, Evarts E. 1984. Tonically discharging putamen neurons exhibit set-dependent responses. *Proc. Natl. Acad. Sci. USA* 81:4998–5001

Kobayashi S, Lauwereyns J, Koizumi M, Sakagami M, Hikosaka O. 2002. Influence of reward expectation on visuospatial processing in macaque lateral prefrontal cortex. *J. Neurophysiol.* 87:1488–98

Lauwereyns J, Watanabe K, Coe B, Hikosaka O. 2002. A neural correlate of response bias in monkey caudate nucleus. *Nature* 418:413–17

Leon MI, Shadlen MN. 1999. Effect of expected reward magnitude on the responses of neurons in the dorsolateral prefrontal cortex of the macaque. *Neuron* 24:415–25

Liu Z, Richmond BJ. 2000. Response differences in monkey TE and perirhinal cortex: stimulus association related to reward schedules. *J. Neurophysiol.* 83:1677–92

Ljungberg T, Apicella P, Schultz W. 1992. Responses of monkey dopamine neurons during learning of behavioral reactions. *J. Neurophysiol.* 67:145–63

Mackintosh NJ. 1975. A theory of attention: variations in the associability of stimulus with reinforcement. *Psychol. Rev.* 82:276–98

Markowitsch HJ, Pritzel M. 1976. Reward-related neurons in cat association cortex. *Brain Res.* 111:185–88

Matsumoto K, Suzuki W, Tanaka K. 2003. Neuronal correlates of goal-based motor selection in the prefrontal cortex. *Science* 301: 229–32

Maunsell JHR. 2004. Neuronal representations of cognitive state: reward or attention? *Trends Cogn. Sci.* 8:261–65

McCoy AN, Crowley JC, Haghighian G, Dean HL, Platt ML. 2003. Saccade reward signals in posterior cingulate cortex. *Neuron* 40: 1031–40

Mirenowicz J, Schultz W. 1994. Importance of unpredictability for reward responses in primate dopamine neurons. *J. Neurophysiol.* 72:1024–27

Mirenowicz J, Schultz W. 1996. Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* 379:449–51

Morris G, Arkadir D, Nevet A, Vaadia E, Bergman H. 2004. Coincident but distinct messages of midbrain dopamine and striatal tonically active neurons. *Neuron* 43:133–43

Musallam S, Corneil BD, Greger B, Scherberger H, Andersen RA. 2004. Cognitive control signals for neural prosthetics. *Science* 305:258–62

Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O. 2004. Dopamine neurons can represent context-dependent prediction error. *Neuron* 41:269–80

Nakamura K, Mikami A, Kubota K. 1992. Activity of single neurons in the monkey amygdala during performance of a visual discrimination task. *J. Neurophysiol.* 67:1447–63

Nishijo H, Ono T, Nishino H. 1988. Single neuron responses in amygdala of alert monkey during complex sensory stimulation with affective significance. *J. Neurosci.* 8:3570–83

O'Doherty JP. 2004. Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr. Opin. Neurobiol.* 14:769–76

Pavlov PI. 1927. *Conditioned Reflexes*. London: Oxford Univ. Press

Pearce JM, Hall G. 1980. A model for Pavlovian conditioning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychol. Rev.* 87:532–52

Platt ML, Glimcher PW. 1999. Neural correlates of decision variables in parietal cortex. *Nature* 400:233–38

Pratt WE, Mizumori SJY. 1998. Characteristics of basolateral amygdala neuronal firing on a spatial memory task involving differential reward. *Behav. Neurosci.* 112:554–70

Pratt WE, Mizumori SJY. 2001. Neurons in rat medial prefrontal cortex show anticipatory rate changes to predictable differential rewards in a spatial memory task. *Behav. Brain Res.* 123:165–83

Rao SC, Rainer G, Miller EK. 1997. Integration of what and where in the primate prefrontal cortex. *Science* 276:821–24

Ravel S, Legallet E, Apicella P. 1999. Tonically active neurons in the monkey striatum do not preferentially respond to appetitive stimuli. *Exp. Brain Res.* 128:531–34

Rescorla RA, Wagner AR. 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In *Classical Conditioning II: Current Research and Theory*, ed. AH Black, WF Prokasy, pp. 64–99. New York: Appleton-Century-Crofts

Roesch MR, Olson CR. 2004. Neuronal activity related to reward value and motivation in primate frontal cortex. *Science* 304:307–10

Rogers QR, Harper AE. 1970. Selection of a solution containing histidine by rats fed a histidine-imbalanced diet. *J. Comp. Physiol. Psychol.* 72:66–71

Roitman MF, Wheeler RA, Carelli RM. 2005. Nucleus accumbens neurons are innately tuned for rewarding and aversive taste stimuli, encode their predictors, and are linked to motor output. *Neuron* 45:587–97

Rolls ET, Critchley HD, Mason R, Wakeman EA. 1996. Orbitofrontal cortex neurons: role in olfactory and visual association learning. *J. Neurophysiol.* 75:1970–81

Rolls ET, Scott TR, Sienkiewicz ZJ, Yaxley S. 1988. The responsiveness of neurones in the frontal opercular gustatory cortex of the macaque monkey is independent of hunger. *J. Physiol.* 397:1–12

Rolls ET, Sienkiewicz ZJ, Yaxley S. 1989. Hunger modulates the responses to gustatory stimuli of single neurons in the caudolateral orbitofrontal cortex of the macaque monkey. *Eur. J. Neurosci.* 1:53–60

Satoh T, Nakai S, Sato T, Kimura M. 2003. Correlated coding of motivation and outcome of decision by dopamine neurons. *J. Neurosci.* 23:9913–23

Savage LJ. 1954. *The Foundations of Statistics*. New York: Wiley

Schoenbaum G, Chiba AA, Gallagher M. 1998. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat. Neurosci.* 1:155–59

Schoenbaum G, Chiba AA, Gallagher M. 1999.

Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *J. Neurosci.* 19:1876–84

Schultz W, Apicella P, Scarnati E, Ljungberg T. 1992. Neuronal activity in monkey ventral striatum related to the expectation of reward. *J. Neurosci.* 12:4595–10

Schultz W, Dayan P, Montague RR. 1997. A neural substrate of prediction and reward. *Science* 275:1593–99

Schultz W, Dickinson A. 2000. Neuronal coding of prediction errors. *Annu. Rev. Neurosci.* 23:473–500

Schultz W, Romo R. 1987. Responses of nigrostriatal dopamine neurons to high intensity somatosensory stimulation in the anesthetized monkey. *J. Neurophysiol.* 57:201–17

Shidara M, Aigner TG, Richmond BJ. 1998. Neuronal signals in the monkey ventral striatum related to progress through a predictable series of trials. *J. Neurosci.* 18:2613–25

Shidara M, Richmond BJ. 2002. Anterior cingulate: single neuron signals related to degree of reward expectancy. *Science* 296:1709–11

Stevens SS. 1957. On the psychophysical law. *Psychol. Rev.* 64:153–81

Sugrue LP, Corrado GS, Newsome WT. 2004. Matching behavior and the representation of value in the parietal cortex. *Science* 304:1782–87

Thorndike EL. 1911. *Animal Intelligence: Experimental Studies*. New York: MacMillan

Thorpe SJ, Rolls ET, Maddison S. 1983. The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp. Brain Res.* 49:93–115

Tobler PN, Dickinson A, Schultz W. 2003. Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. *J. Neurosci.* 23:10402–10

Tobler PN, Fiorillo CD, Schultz W. 2005. Adaptive coding of reward value by dopamine neurons. *Science* 307:1642–45

Tremblay L, Hollerman JR, Schultz W. 1998. Modifications of reward expectation-related neuronal activity during learning in primate striatum. *J. Neurophysiol.* 80:964–77

Tremblay L, Schultz W. 1999. Relative reward preference in primate orbitofrontal cortex. *Nature* 398:704–8

Tremblay L, Schultz W. 2000a. Reward-related neuronal activity during go-nogo task performance in primate orbitofrontal cortex. *J. Neurophysiol.* 83:1864–76

Tremblay L, Schultz W. 2000b. Modifications of reward expectation-related neuronal activity during learning in primate orbitofrontal cortex. *J. Neurophysiol.* 83:1877–85

Ungless MA, Magill PJ, Bolam JP. 2004. Uniform inhibition of dopamine neurons in the ventral tegmental area by aversive stimuli. *Science* 303:2040–42

von Neumann J, Morgenstern O. 1944. *The Theory of Games and Economic Behavior*. Princeton, NJ: Princeton Univ. Press

Waelti P. 2000. *Activité phasique des neurones dopaminergiques durant une tâche de discrimination et une tâche de blocage chez le primate vigile*. PhD thesis. Univ. de Fribourg, Switzerland

Waelti P, Dickinson A, Schultz W. 2001. Dopamine responses comply with basic assumptions of formal learning theory. *Nature* 412:43–48

Wallis JD, Miller EK. 2003. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.* 18:2069–81

Wang Y, Cummings SL, Gietzen DW. 1996. Temporal-spatial pattern of c-fos expression in the rat brain in response to indispensable amino acid deficiency. I. The initial recognition phase. *Mol. Brain Res.* 40:27–34

Watanabe M. 1989. The appropriateness of behavioral responses coded in post-trial activity of primate prefrontal units. *Neurosci. Lett.* 101:113–17

Watanabe M. 1996. Reward expectancy in primate prefrontal neurons. *Nature* 382:629–32

Watanabe M, Hikosaka K, Sakagami M, Shirakawa SI. 2002. Coding and monitoring of behavioral context in the primate prefrontal cortex. *J. Neurosci.* 22:2391–400

Weber EH. 1850. Der Tastsinn und das Gemeingefuehl. In *Handwoerterbuch der Physiologie*, Vol. 3, Part 2, ed. R Wagner, pp. 481–588. Braunschweig, Germany: Vieweg

Wise RA. 2002. Brain reward circuitry: insights from unsensed incentives. *Neuron* 36:229–40

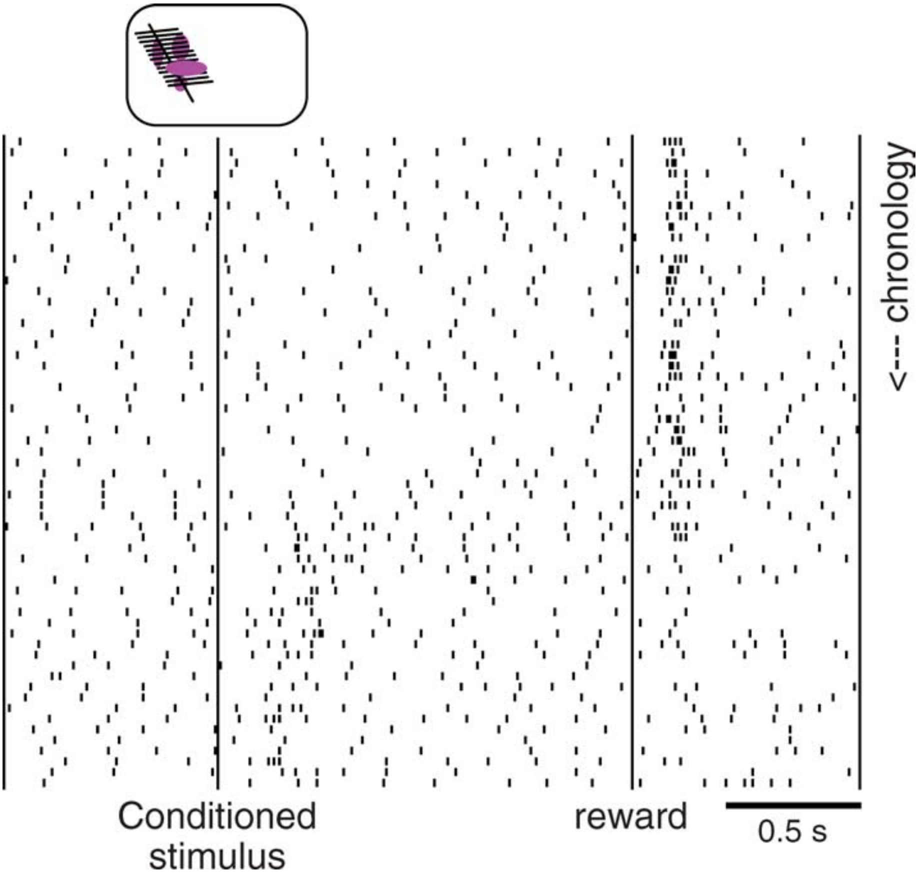


Figure 2 Testing the contiguity requirement for associative learning: acquisition of neural response in a single dopamine neuron during a full learning episode. Each line of dots represents a trial, each dot represents the time of the discharge of the dopamine neuron, the vertical lines indicate the time of the stimulus and juice reward, and the picture above the raster shows the visual conditioned stimulus presented to the monkey on a computer screen. Chronology of trials is from top to bottom. The top trial shows the activity of the neuron while the animal saw the stimulus for the first time in its life, whereas it had previous experience with the liquid reward. Data from Waelti (2000).

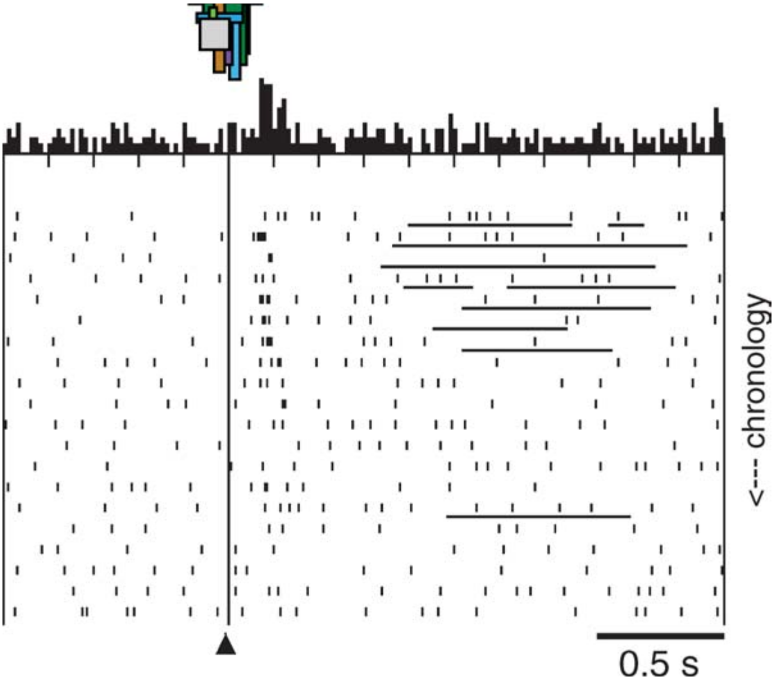
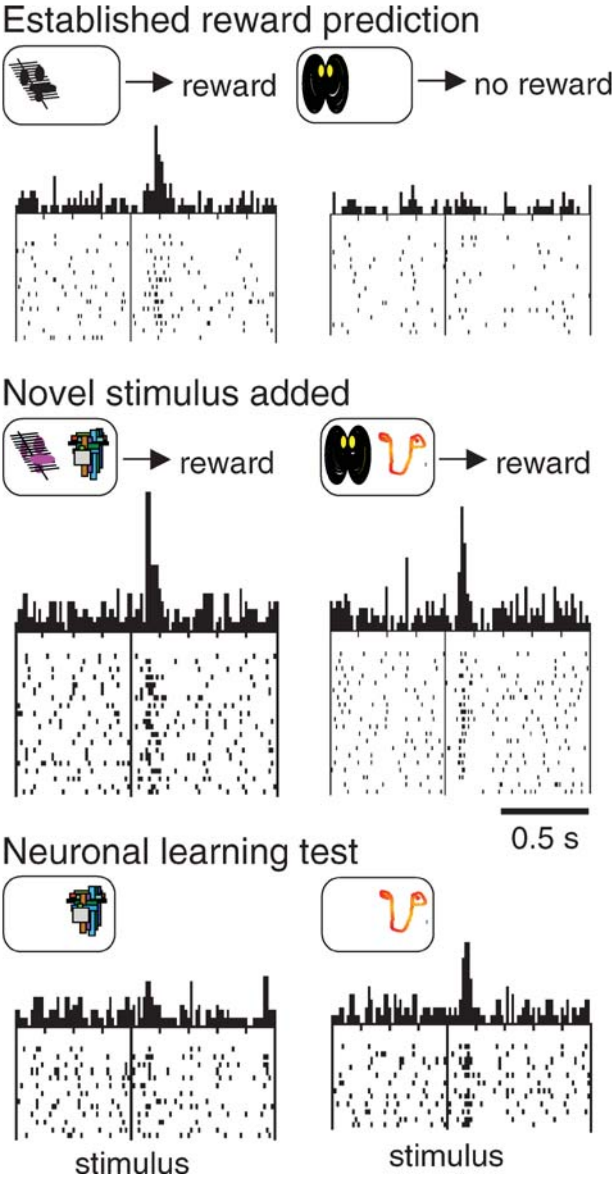


Figure 5 Loss of response in dopamine neuron to the conditioned stimulus following withholding of reward. This manipulation violates the contiguity requirement (co-occurrence of reward and stimulus) and produces a negative prediction error that brings down the associative strength of the stimulus. The contingency moves toward the neutral situation. Data from Tobler et al. (2003).



See legend on next page

Figure 6 Acquisition of dopamine response to reward-predicting stimulus is governed by prediction error. Neural learning is blocked when the reward is predicted by another stimulus (*left*) but is intact in the same neuron when reward is unpredicted in control trials with different stimuli (*right*). The neuron has the capacity to respond to reward-predicting stimuli (*top left*) and discriminates against unrewarded stimuli (*top right*). The addition of a second stimulus results in maintenance and acquisition of response, respectively (*middle*). Testing the added stimulus reveals absence of learning when the reward is already predicted by a previously conditioned stimulus (*bottom left*). Data from Waelti et al. (2001).

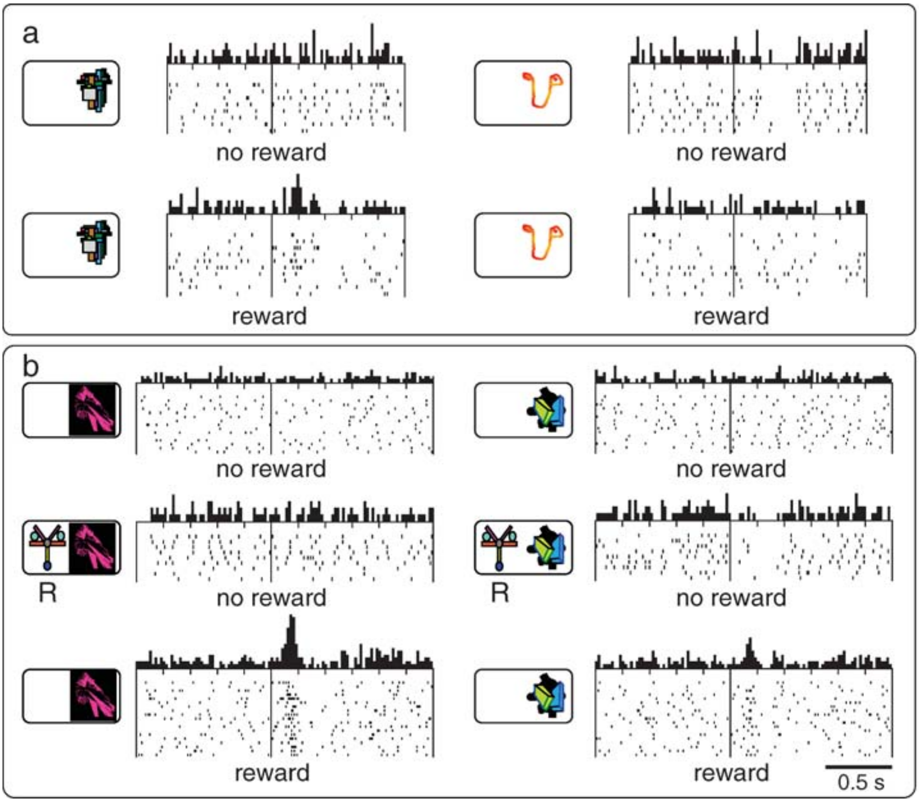


Figure 8 Coding of prediction errors by dopamine neurons in specific paradigms. (*a*) Blocking test. Lack of response to absence of reward following the blocked stimulus, but positive signal to delivery of reward (*left*), in contrast to control trials with a learned stimulus (*right*). Data from Waelti et al. 2001. (*b*) Conditioned inhibition task. Lack of response to absence of reward following the stimulus predicting no reward (*top*), even if the stimulus is paired with an otherwise reward-predicting stimulus (R, *middle*, summation test), but strong activation to reward following a stimulus predicting no reward (*bottom*). These responses contrast with those following the neutral control stimulus (*right*). Data from Tobler et al. (2003).

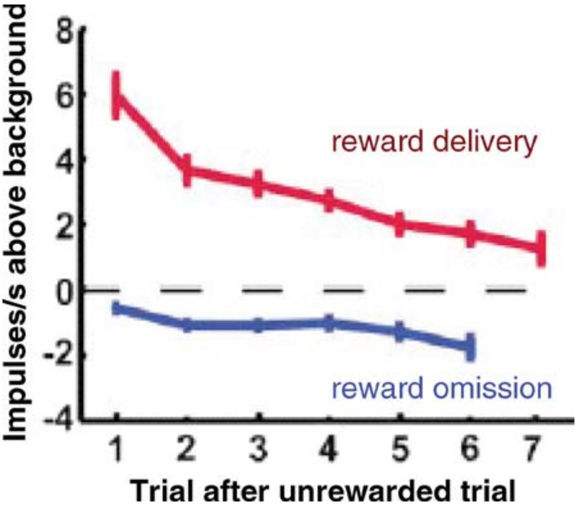


Figure 9 Time information contained in predictions acting on dopamine neurons. In the particular behavioral task, the probability of reward, and thus the reward prediction, increases with increasing numbers of trials after the last reward, reaching p = 1.0 after six unrewarded trials. Accordingly, the positive dopamine error response to a rewarding event decreases over consecutive trials (*upper curve*), and the negative response to a nonrewarding event becomes more prominent (*lower curve*). Data are averaged from 32 dopamine neurons studied by Nakahara et al. (2004), © Cell. Press.

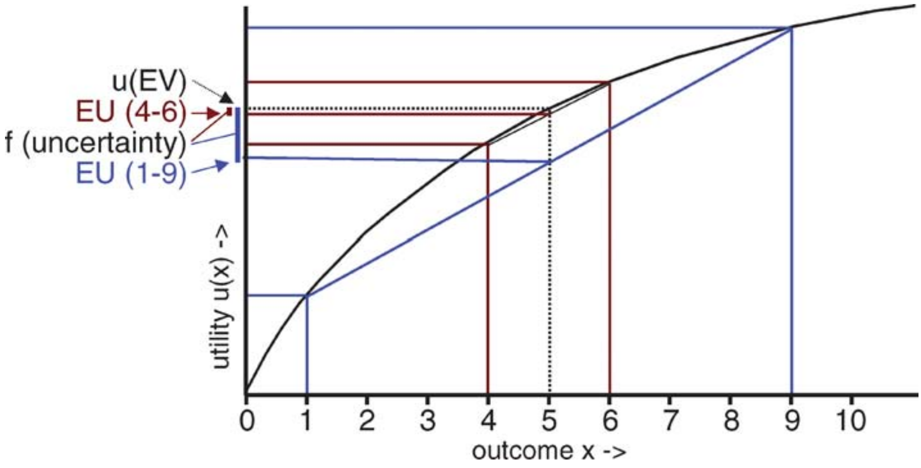


Figure 12 A hypothetical concave utility function. EV, expected value (5 in both gambles with outcomes of 1 and 9, and 4 and 6); EU, expected utility. See text for description.

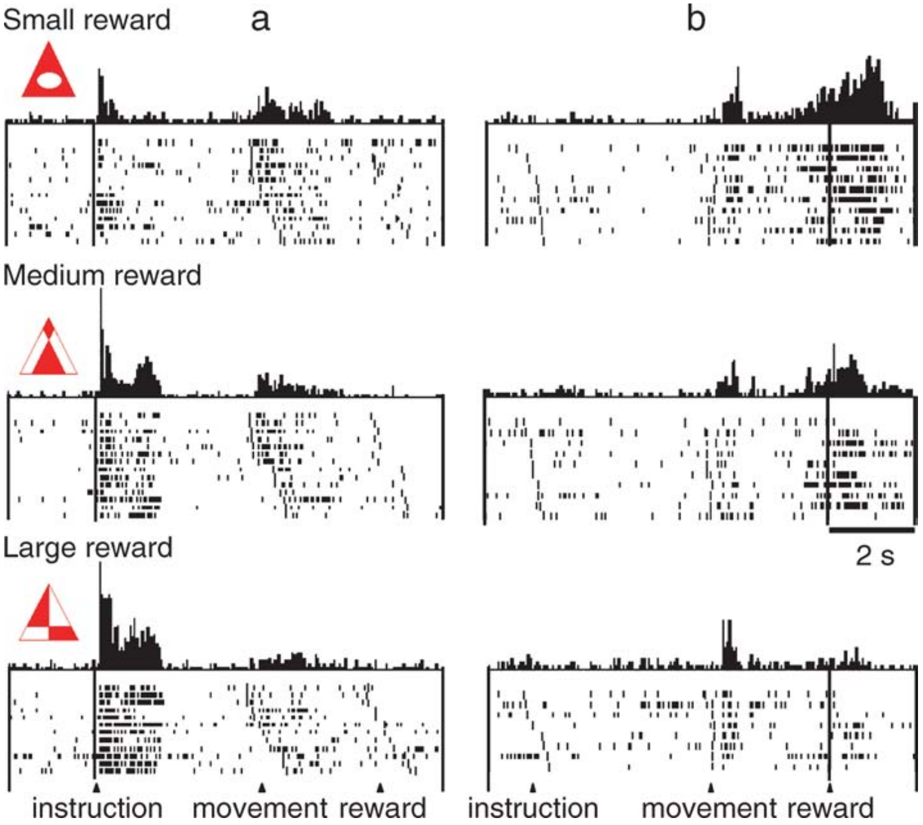


Figure 13 Discrimination of reward magnitude by striatal neurons. (*a*) Increasing response in a caudate neuron to instruction cues predicting increasing magnitudes of reward (0.12, 0.18, 0.24 ml). (*b*) Decreasing response in a ventral striatum neuron to rewards with increasing volumes. Data from Cromwell & Schultz 2003.

Annual Review of Psychology
Volume 57, 2006

Contents

Frontispiece—*Herbert C. Kelman* xvi

Prefatory

Interests, Relationships, Identities: Three Central Issues for Individuals and Groups in Negotiating Their Social Environment, *Herbert C. Kelman* 1

Brain Mechanisms and Behavior: Emotion and Motivation

Emotion and Cognition: Insights from Studies of the Human Amygdala, *Elizabeth A. Phelps* 27

Stress and Neuroendocrinology

Stressful Experience and Learning Across the Lifespan, *Tracey J. Shors* 55

Reward and Addiction

Behavioral Theories and the Neurophysiology of Reward, *Wolfram Schultz* 87

Genetics of Behavior

Genetics of Affective and Anxiety Disorders, *E.D. Leonardo and René Hen* 117

Sleep

Sleep, Memory, and Plasticity, *Matthew P. Walker and Robert Stickgold* 139

Comparative Psychology, Ethology, and Evolution

Neuroecology, *David F. Sherry* 167

Evolutionary Psychology

The Evolutionary Psychology of Facial Beauty, *Gillian Rhodes* 199

Language and Communication

Explanation and Understanding, *Frank C. Keil* 227

Adolescence

Adolescent Development in Interpersonal and Societal Contexts, *Judith G. Smetana, Nicole Campione-Barr, and Aaron Metzger* 255

Individual Treatment

Enduring Effects for Cognitive Therapy in the Treatment of Depression and Anxiety, *Steven D. Hollon, Michael O. Stewart, and Daniel Strunk* 285

FAMILY/MARITAL THERAPY

Current Status and Future Directions in Couple Therapy, *Douglas K. Snyder, Angela M. Castellani, and Mark A. Whisman* 317

ATTITUDE CHANGE AND PERSUASION

Attitudes and Persuasion, *William D. Crano and Radmila Prislin* 345

BARGAINING, NEGOTIATION, CONFLICT, SOCIAL JUSTICE

Psychological Perspectives on Legitimacy and Legitimation, *Tom R. Tyler* 375

INDIVIDUAL DIFFERENCES AND ASSESSMENT

Personality and the Prediction of Consequential Outcomes, *Daniel J. Ozer and Verónica Benet-Martínez* 401

ENVIRONMENTAL PSYCHOLOGY

Child Development and the Physical Environment, *Gary W. Evans* 423

MARKETING AND CONSUMER BEHAVIOR

Consumer Psychology: Categorization, Inferences, Affect, and Persuasion, *Barbara Loken* 453

STRUCTURES AND GOALS OF EDUCATIONAL SETTINGS

Classroom Goal Structure, Student Motivation, and Academic Achievement, *Judith L. Meece, Eric M. Anderman, and Lynley H. Anderman* 487

DATA ANALYSIS

Analysis of Longitudinal Data: The Integration of Theoretical Model, Temporal Design, and Statistical Model, *Linda M. Collins* 505

TIMELY TOPICS

The Internet as Psychological Laboratory, *Linda J. Skitka and Edward G. Sargis* 529

Family Violence, *Patrick Tolan, Deborah Gorman-Smith, and David Henry* 557

Understanding Affirmative Action, *Faye J. Crosby, Aarti Iyer, and Sirinda Sincharoen* 585

INDEXES

Subject Index 613

Cumulative Index of Contributing Authors, Volumes 47–57 637

Cumulative Index of Chapter Titles, Volumes 47–57 642

ERRATA

An online log of corrections to *Annual Review of Psychology* chapters may be found at http://psych.annualreviews.org/errata.shtml