

the cerebellum. Trace eyeblink conditioning also depends on the hippocampus, but this relationship may reflect contributions of declarative memory. Operant conditioning (instrumental learning) is a change in the probability of a behavioral response associated with a reward (or punishment). Instrumental behaviors include flexible goal-directed actions and inflexible stimulus-driven habits. Whereas goal-directed actions depend more on the dorsomedial striatum (caudate) and the hippocampus, stimulus-driven habits are more dependent on the dorsolateral striatum (putamen).

Cellular Mechanisms of Memory

The preceding sections reviewed the major neural substrates of memory at the *system level*—that is, the extent to which memory depends on gross anatomical structures such as the medial temporal lobes, the cortex, the basal ganglia, and the cerebellum. At the system level the neural substrates of memory are clearly heterogeneous. At the *cellular and molecular levels*, however, *all* the forms of memory seem to depend on changes in neural connectivity and the relative strength of synaptic transmission.

Early in the twentieth century, the German evolutionary biologist Richard Semon coined the term **engram** to refer to the physical and biochemical changes underlying memory storage in the brain. Two related questions about engrams, or **memory traces**, have long fascinated neuroscientists: where are they located in the brain, and how, in neurobiological terms, are they formed?

With respect to the where question, an early idea inspired by Pavlovian conditioning was that engrams depend on the integrity of the pathways connecting sensory and motor regions. In the 1920s, American neuroscientist Karl Lashley tested this idea by making cuts in a rat's cortex that would disconnect sensory and motor regions before or after the rats learned to run mazes. Lashley found that the location of the lesions did not matter much, and that only the extent of the tissue destroyed was consequential. This finding is consistent with the idea that an engram consists of a network of neurons with connections distributed over the cortex; thus, only when a large proportion of the network is destroyed is the engram seriously compromised (see the discussion of graceful degradation in **Box 8C**). This idea does not imply that engrams are randomly distributed. As mentioned already, the modern view is that engrams are defined primarily by the connectivity of the brain regions originally involved in processing the relevant category of information. That is, memory traces for visual information are stored mostly in visual cortices, those for auditory information in auditory cortices, those for motor skills in the motor cortices, and so on.

This general answer to the question of where engrams are located suggests an answer to *how* they are formed. If engrams consist of networks of neurons, then they are formed by a process whereby neurons that once functioned relatively independently begin to work together by virtue of strengthened connections between them. This broad answer was first explicitly proposed in 1949 by Canadian psychologist Donald Hebb, who hypothesized that memories are stored in the brain in the form of networks of neurons that he called *cell assemblies*. Hebb's idea was that through experience (i.e., learning), assemblies gradually become specifically associated with objects, actions, or concepts. To explain how the relevant neurons come to be linked, he proposed that when presynaptic and postsynaptic neurons are simultaneously active (i.e., fire action potentials at the same time), the synaptic connections between them are strengthened. As a result, synaptic associations grow stronger and tend to persist. This mechanism, known today as **Hebbian learning**, is captured in the adage “cells that fire together wire together.”

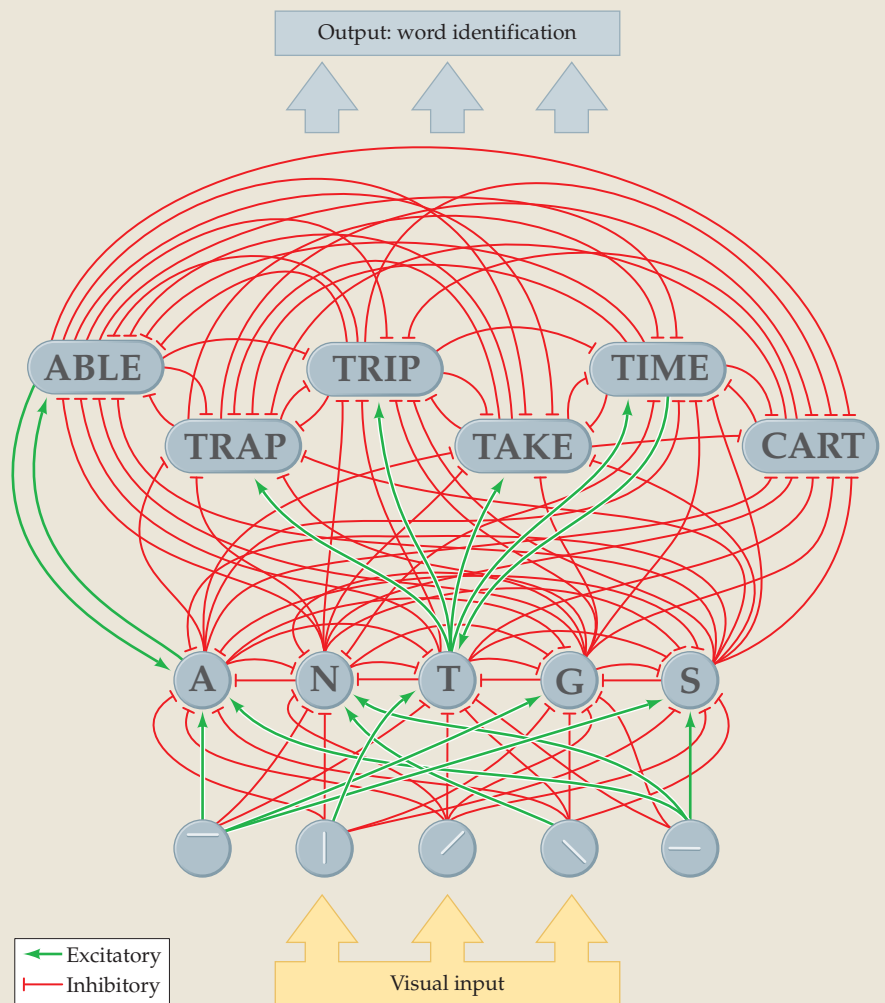
BOX 8C CONNECTIONIST MODELS

Unlike sequential processing models, connectionist models assume that information processing occurs in parallel across a large number of distributed units. For this reason, these models are also known as *parallel distributed processing (PDP) models*. The units, or *nodes*, in connectionist schemes are linked in a network, much as neurons are interconnected in the brain. Thus, connectionist models are also known as *neural network models*.

An example is the model of written-word recognition developed by cognitive scientists James McClelland and his colleague David Rumelhart. Like other connectionist models, it is a network consisting of layers of units. The units of the *input layer* receive information from the environment; in this example, each input unit receives visual information and responds to one particular feature of a written letter (e.g., a horizontal line). The *output layer* consists of units that indicate the outcomes of the processing within the network. Given that this is a model of word recognition, the output units correspond to all the words the model can recognize. Finally, input and output units are connected through one or more *hidden layers*, so called because their units are invisible from the outside. For simplicity, the model in the example here shows only a few letter features, letters, and words.

The network shown responds to only the *first letter* of each word; similar networks would be required to recognize each of the other letters. The connections between units can be excitatory or inhibitory. Some excitatory and inhibitory connections in the figure are obvious. For example, the letter *a* excites the word *able* but inhibits the word *trap* because *able* starts with an *a* but *trap* does not. The inhibitory connections among the units within the same layer ensure that one option will tend to win out over the others. Activation flows not only from input to output layers but also from output to input layers (e.g., the excitatory connection from *able* to the letter *a*).

To see how the model works, consider the sentence "John packed his suitcase and left for a trip." Before such a network reaches the end of the sentence, the node for the word *trip* is activated by semantic associations from the words *suitcase* and *left* (not shown in the figure). This activation facilitates the identification of the letters *t*, *r*, *i*, and *p* and thus the corresponding feature detectors. At the



A connectionist model for written-word recognition. Note that the feature and letter units represent only the first position in each four-letter word. (After McClelland and Rumelhart 1981.)

same time, activation flows in the opposite direction from feature to letter to word detectors in the input layer. A strength of models like this is that they specify how conceptually driven processing (already established in the network) and data-driven processing might interact. Another strength is that they entail a built-in learning mechanism: the modification of connection weights represents stored information that alters the output.

In computer simulations, networks are often "trained" using *supervised learning algorithms* such as *back-propagation*, in which the actual output (e.g., *trap*) is compared to the desired output (e.g., *trip*) and the weights of the connections from the layers below are modified to

reduce subsequent errors. Learning can also be *unsupervised*; for example, the strength of the connection between units may be automatically increased whenever the units are simultaneously active (Hebbian learning, described in the text).

PDP models are attractive in cognitive neuroscience because changing the strength of synaptic connections through individual experience or natural selection is clearly the major way nervous systems store information. The assumption that knowledge is not stored in a single location but is distributed across numerous sites is also consistent with much neu-

(Continued on next page)

■ BOX 8C (continued)

robiological evidence. Local damage to either the brain or a connectionist model typically degrades performance but doesn't abolish it altogether—a phenomenon called *graceful degradation*.

At the same time, connectionist models have several limitations, including difficulty accounting for rule-based processes in human reasoning; difficulty discerning the strategies being used to solve problems by examining weight patterns of the trained network architecture; and, consequently, difficulty distinguishing among alternative models of the same cognitive process. These caveats notwithstanding, connectionist models

have provided cognitive scientists with important insights and a powerful tool.

References

- HAYKIN, S. (1998) *Neural Networks: A Comprehensive Foundation*, 2nd Ed. New York: Prentice-Hall.
- MCCLELLAND, J. L. AND D. E. RUMELHART (1981) An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychol. Rev.* 88: 375–407.
- MCCLELLAND, J. L., D. E. RUMELHART AND G. E. HINTON (2002) The appeal of parallel distributed processing. In *Foundations of Cognitive Psychology: Core Reading*, D. J. Levitin (ed.). Cambridge, MA: MIT Press, pp. 57–91.
- MCCLELLAND, J. L., D. E. RUMELHART AND THE PDP RESEARCH GROUP (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 2: *Psychological and Biological Models*. Cambridge, MA: MIT Press.
- RUMELHART, D. E., J. L. MCCLELLAND AND THE PDP RESEARCH GROUP (1986) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Vol. 1: *Foundations*. Cambridge, MA: MIT Press.

Even though they were articulated more than 50 years ago, Hebb's proposals, in essence, remain a plausible basis for understanding memory in cellular terms. Only in the last few decades, however, have neuroscientists begun to understand some of the cellular and molecular mechanisms that allow a strengthening of synaptic connectivity as a result of coincident neuronal activity.

Habituation and sensitization

One of the main difficulties in investigating the cellular bases of memory is the sheer complexity of neuronal circuits and behaviors, particularly in the mammalian brain. One way to simplify this problem is to investigate memory mechanisms in an organism with fewer neurons and a more limited behavioral repertoire. Neuroscientist Eric Kandel and his colleagues pioneered this approach when, beginning in the late 1960s, they carried out studies using the sea slug *Aplysia californica*. The ganglia that make up the nervous system in this animal contain only a few thousand neurons, many of which are large and individually identifiable. Despite the simplicity of its nervous system, the sea slug shows rudimentary learning abilities, which Kandel and his group measured in terms of changes in a simple withdrawal reflex.

A sea slug withdraws its gill when its siphon is lightly touched, presumably as a defensive maneuver, and this simple reflex demonstrates two forms of learning that occur in many animals, including humans: habituation and sensitization. **Habituation** is a reduced response when the same stimulus is repeated over and over; **sensitization** is an increased response to the habituated stimulus when it is paired with an aversive stimulus such as a shock to the animal's tail (**Figure 8.23**). Touching the siphon skin stimulates sensory neurons, which in turn excite interneurons and motor neurons; the excited motor neurons then elicit withdrawal of the gill.

Electrical recordings show that habituation involves a decrease in neurotransmitter release at the synapses between sensory neurons and motor neurons, whereas sensitization involves an increase in neurotransmitter release at these synapses. Whereas the functional changes underlying habituation appear to be confined to the sensory neuron–motor neuron synapse, the tail shock used to elicit sensitization recruits a wider array of sensory neurons, which in turn excite modulatory interneurons that increase neurotransmitter release from the siphon sensory neurons, enhancing gill withdrawal.

Both habituation and sensitization are simple forms of memory. In particular, the modulatory effect of tail shock on the gill withdrawal reflex lasts a matter of minutes, and it can thus be regarded as a model for short-term memory in this system. Moreover, repeated tail shocks over longer periods trigger gene expression, new protein synthesis, and the formation of new synaptic connections, all of which result in enhancement of the gill withdrawal reflex that can last weeks—a model of simple long-term memory.

Long-term potentiation and depression

A second major advance in understanding the cellular mechanisms of memory was accomplished in the early 1970s in studies of hippocampal synaptic transmission carried out by neuroscientists Terje Lømo and Timothy Bliss. As illustrated in **Figure 8.24**, the researchers stimulated different afferent pathways (e.g., pathways 1 and 2) and recorded responses in a postsynaptic neuron (postsynaptic potentials). They found that when one pathway (e.g., pathway 1) was stimulated with a high-frequency electrical train, the postsynaptic neuron later showed stronger responses to input from the stimulated pathway but not from other pathways (e.g., pathway 2). Because this focal enhancement lasted a long time, they called the phenomenon **long-term potentiation (LTP)**. Although LTP was first localized in the rabbit hippocampus, subsequent research showed that it occurs in many other species and brain regions, including the cortex, the amygdala, the basal ganglia, and the cerebellum. Depending on the locus and stimulation paradigm, LTP can last tens of minutes, hours, or longer.

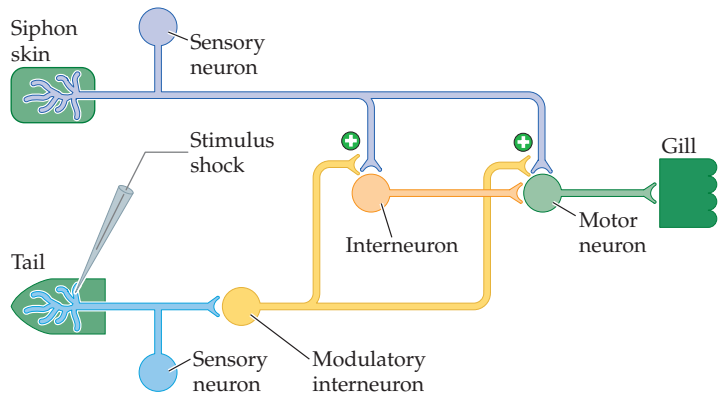


Figure 8.23 Two forms of learning in the sea slug *Aplysia* The simple neuronal circuit in the abdominal ganglion of the sea slug is used in studies of habituation and sensitization. Touching the animal's siphon elicits withdrawal of the gill. When the siphon is repeatedly touched, this response progressively decreases (habituation). Conversely, a shock to the tail enhances the habituated gill withdrawal response (sensitization). (After Squire and Kandel 1999.)

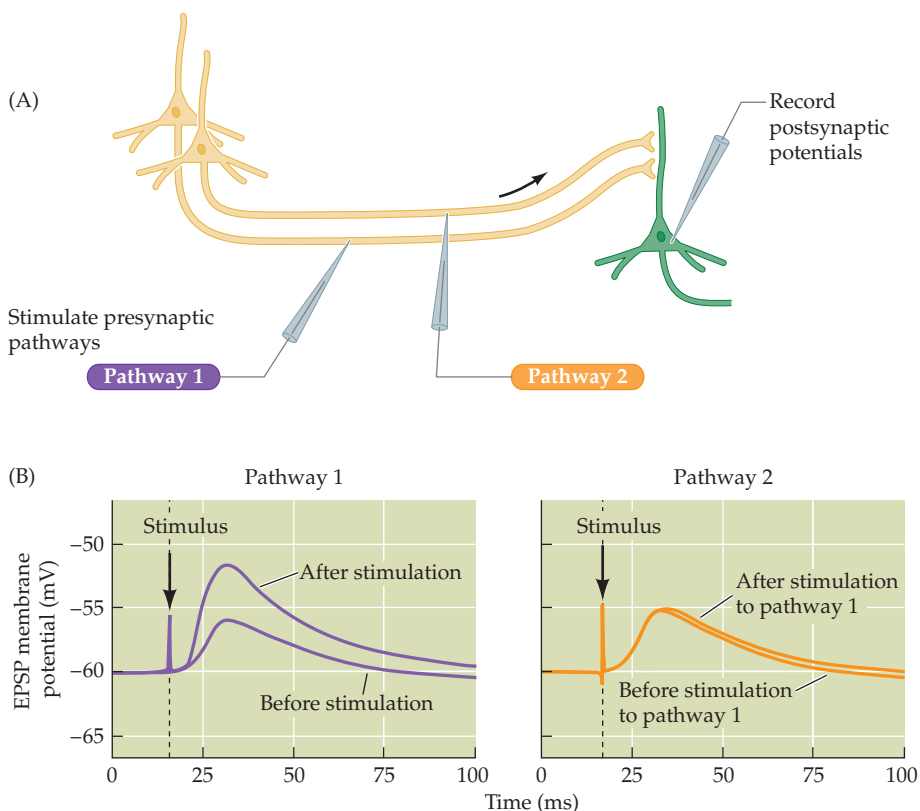


Figure 8.24 Long-term potentiation (LTP) (A) Two pathways in the hippocampus that can be stimulated independently. (B) Following high-frequency stimulation (not shown), the postsynaptic potential elicited by a single stimulus is enhanced, but only for the stimulated pathway. EPSP, excitatory postsynaptic potential. (After Malinow et al. 1989.)

LTP has several properties that advance it as an attractive candidate for one of the major cellular mechanisms of memory. First, LTP can be induced by a single high-frequency train of stimulation—a feature that could explain how some memories can be formed by a single experience. Second, the fact that LTP can last for days or even weeks could explain the persistence of memory over these sorts of time intervals. Third, LTP has a degree of **specificity** that fits well with the specificity of memories. Only those synapses activated during stimulation are enhanced; other synapses—even other synapses on the same neuron—are not affected. Finally, LTP has a property of **associativity** that could underlie memory mechanisms such as Hebbian learning: if a weak and a strong pathway impinging on a target neuron were activated at the same time, then both pathways would tend to become associated via the output of the target cell.

As suggested by the phenomenon of habituation, synaptic *weakening* is equally a part of learning and memory. If synapses continued to increase in strength because of LTP or a similar mechanism, they would eventually reach a point of saturation. It makes sense, therefore, for synaptic potentiation to be counterbalanced by a mechanism of synaptic inhibition. Such a mechanism, known as **long-term depression (LTD)**, was found in the late 1970s. Whereas LTP occurs with high-frequency stimulation over a brief period, LTD is elicited by low-frequency stimulation over a longer period. LTD can also last several hours and is again specific to activated synapses. As might be expected, LTD can counteract LTP, and vice versa. There is also some evidence that LTD can enhance LTP in neighboring synapses and that it can block the formation of associations that do not follow Hebb's rule.

Linking LTP to memory performance

Although LTP has many of the properties required for memory storage and behavior, it has been studied mainly *in vitro*. Thus, directly linking LTP and memory performance has been a major research effort in recent years.

One approach has been to seek evidence of **behavioral LTP**—that is, a demonstrated change in synaptic efficacy similar to LTP that follows a natural learning experience. For example, rats raised in an enriched stimulatory environment show enhanced postsynaptic potentials when compared to rats raised in an environment impoverished in stimulation and learning experiences. The enhancement disappears within a few weeks if the enriched rats are placed in the impoverished environment, suggesting that enriched experience is necessary to maintain strong synapses. In addition, changes in synaptic efficacy lasting about 30 minutes have been observed in the rat hippocampus following exploration of a novel environment, which could be indicative of learning and memory storage. There is also evidence that inducing LTP in sensory pathways can enhance synaptic responses to natural sensory stimulation, and that learning experiences can lead to long-term increases in correlated firing among hippocampal neurons that were active during learning (thus showing associativity).

Another approach to linking LTP to naturally occurring learning and memory has been to *prevent* LTP and then determine whether memory is impaired. One way to disrupt LTP is to use “knockout” mice in which a gene for a protein crucial to LTP is eliminated by gene splicing techniques. Neuroscientist Susumu Tonegawa and his colleagues produced a strain of mice that lacked the gene for NMDA receptors in hippocampal region CA1. This receptor protein mediates certain forms of LTP, and indeed the knockout mice showed little or no LTP in the CA1 region of the hippocampus. Mice lacking NMDA receptors were also severely impaired in spatial learning (a property mediated by the hippocampus in rodents), although they were unimpaired in nonspatial learning. Whereas control mice searched for a platform hidden in a water tank only near the location where the platform had been located in previous trials (the *Morris water*

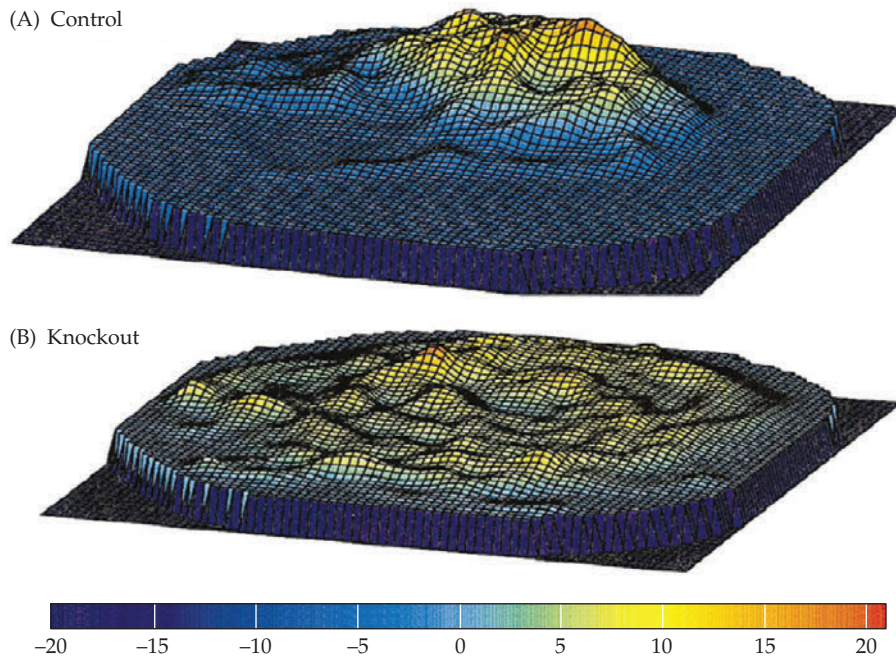


Figure 8.25 Genetic impact on learning the water maze task (A) The landscape shows that control mice focused their search in the trained area, resulting in a high total occupancy peak (yellow) in the old platform position, which they apparently remembered. (B) Laboratory mice in which the gene for NMDA receptors was eliminated (“knocked out”) in the hippocampus navigated more or less randomly over the full area of the pool, apparently failing to have learned where the platform should be located. (From Tsien et al. 1996.)

maze; see Box 8A), the NMDA-deficient mice searched widely for the platform, indicating that they had little spatial memory (**Figure 8.25**).

Efforts to establish behaviorally relevant evidence of LTP involvement have not been limited to the hippocampus or to spatial learning. In the mid-1990s, neuroscientist Joseph LeDoux and collaborators investigated behavioral LTP in the context of auditory fear conditioning, focusing on the pathway connecting the medial geniculate nucleus of the thalamus (a relay station in the auditory sensory pathways; see Chapter 4) to the amygdala. They found that electrical stimulation of this pathway gave rise to LTP in the lateral nucleus of the amygdala, in accord with the fact that synaptic responses in the same nucleus are enhanced by natural auditory stimulation. Although this question is far from settled, there is at least some evidence that LTP is correlated with newly learned behaviors.

Learning-related changes in synaptic morphology

Maintenance of any mechanism of memory that is to last for a long period requires gene expression and the synthesis of new proteins. Among other effects, these changes can, and in some cases do, lead to long-lasting structural alternations in synapses. These morphological changes in neuronal connectivity can be considered the culmination of a process of *synaptic consolidation* discussed in Chapter 9.

For many neurons, learning-related alterations in synaptic morphology appear to involve **dendritic spines**. These are small protrusions from dendritic branches—often mushroom-shaped—that in the neocortex receive excitatory synaptic input from other neurons. Several sorts of morphological changes in dendritic spine structure over periods of a few minutes to hours have been

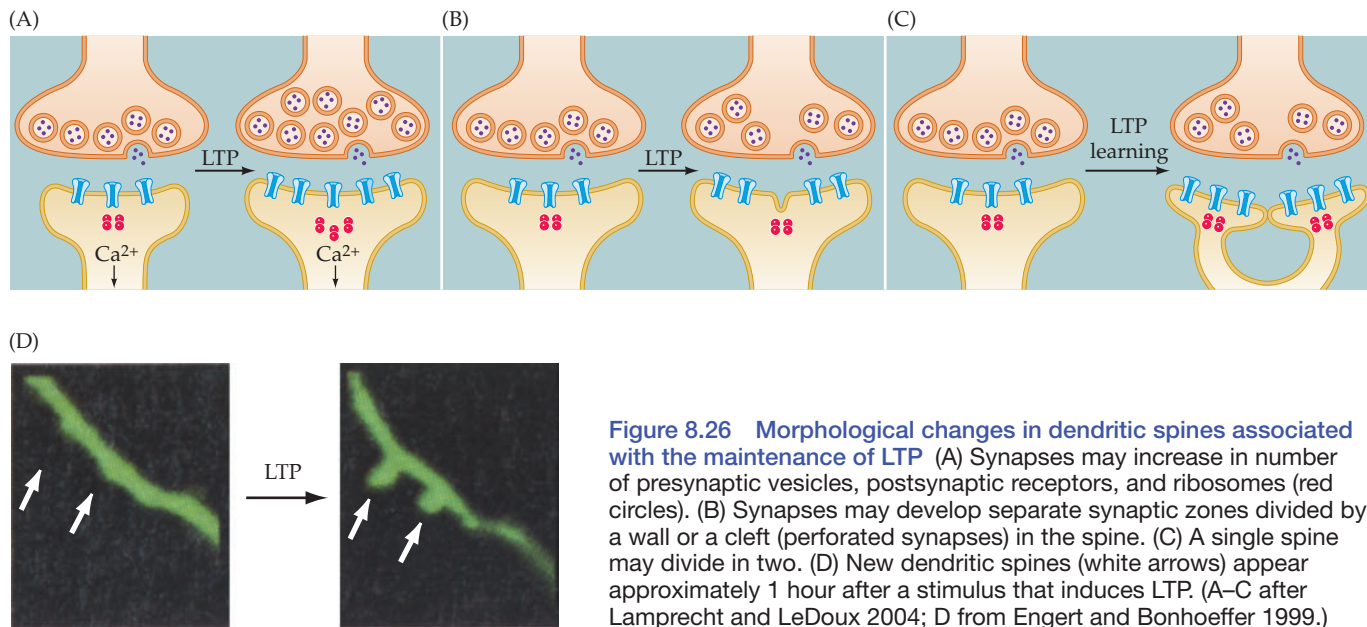


Figure 8.26 Morphological changes in dendritic spines associated with the maintenance of LTP (A) Synapses may increase in number of presynaptic vesicles, postsynaptic receptors, and ribosomes (red circles). (B) Synapses may develop separate synaptic zones divided by a wall or a cleft (perforated synapses) in the spine. (C) A single spine may divide in two. (D) New dendritic spines (white arrows) appear approximately 1 hour after a stimulus that induces LTP. (A–C after Lamprecht and LeDoux 2004; D from Engert and Bonhoeffer 1999.)

observed, some in parallel with the development of LTP and changes in behavior. The simplest changes are in the length and diameter of the “stem” of the mushroom (the spine neck) and/or changes in the size of the mushroom itself (Figure 8.26A). The former changes affect passage of the synaptic signal from the spine to the dendritic shaft, and the latter allow a change in the number of presynaptic terminals in contact with a spine (Figure 8.26B,C). Finally, new spines can emerge from the shaft of the dendrite and others disappear, thereby changing the number of spines and thus synapses in yet another way (Figure 8.26D). All such changes could support memories or their loss by increasing or decreasing synaptic strength over long periods.

In summary, the physical representations of memories (engrams) are assumed to consist of functional and structural changes in synaptic connectivity among neurons that have strengthened or weakened as a result of experience. Long-term maintenance of LTP and other memory mechanisms presumably requires changes in gene expression and the synthesis of new proteins that eventually lead to structural alternations in synaptic morphology and neuronal connectivity. This commonality of mechanisms at the cellular level is quite compatible with functional heterogeneity at the systems level.

Summary

1. The generally accepted taxonomy of memory systems distinguishes three major systems: declarative memory, nondeclarative memory, and working memory. Damage to the medial temporal lobe impairs declarative memory but spares working memory. Other types of lesions can impair working memory without disrupting declarative memory.
2. Nondeclarative memory depends on brain regions other than medial temporal lobes, which vary for different forms of nondeclarative memory: priming, skill learning, and conditioning.
3. Priming is defined as a change in processing of a stimulus due to a previous encounter with the same or a related stimulus in the absence of conscious awareness of the original encounter. The neural correlates of priming include neocortical regions that vary with the type of priming: visual perceptual priming is associated with altered activity in the visual sensory areas in occipitotemporal cortex; conceptual priming, with anterior left inferior prefrontal areas; and semantic priming,