Learning and Memory, Models of

Introductory article

David M Eagleman, Salk Institute, La Jolla, California, USA
P Read Montague, Baylor College of Medicine, Houston, Texas, USA

CONTENTS

Learning is the process of acquiring knowledge, and memory is the retention of knowledge. Both have their roots in the biochemistry and anatomy of the brain, and are commonly studied by investigating the connection between brain cells, the synapse.

INTRODUCTION

Animals interact appropriately with the world through their ability to learn and remember the specifics of their environment. Learning and memory are possible only because of the changeable properties of the vast networks of cells in the brain. In this way, neural activity can dynamically modify the brain's organization. Learning and memory can be studied at many levels – biochemical, cellular and systems – and even by investigation of perturbations that decrease learning and memory, from genetic (Down syndrome) to pathological (Alzheimer disease).

What do we mean by learning and memory, and how do we detect when it happens? The answer seems complicated by the fact that memory is not a monolithic entity, but comprises many different types. Memory can be divided into declarative learning (names, facts) and nondeclarative learning (riding a bicycle), and within these categories are scores of subtypes. What neural processes could underlie such a variety of different types of memory?

Specialized Brain Areas

Examples of brain damage demonstrate that different areas of the brain are involved in different subtypes of learning and memory. For example, injury to the medial temporal lobe of the cerebral cortex affects declarative memory but not nondeclarative memory. The cerebellum, on the other hand, seems to be important in learning certain motor skills, especially those involving balance and coordination. The basal ganglia are evidently important in learning that links rewards with motor activity. The list of brain structures and their involvement with learning and memory is large and ever-increasing, and it is interesting to note that the integrity of a particular subsystem is not always essential to the functioning of others. That is, one can lose the ability to learn dates and facts (as in amnesia), but this has no bearing on the ability to learn and remember new motor skills. Much of what we will discuss here will be distilled from data concerning two main areas: the cortex and hippocampus.

The hippocampus (and its surrounding regions) seems to be a central organ of learning, and its structure makes its physiology very amenable to laboratory study; it is perhaps the best-studied area of the brain. In 1953, a 27-year-old patient referred to as HM had his hippocampus and surrounding areas surgically removed to relieve intractable epilepsy. Thereafter, HM lost his ability to form new memories or learn new facts, and although he could acquire new skills, he had no memory of having acquired them.

General Cellular Principles?

While learning and memory can be different in specialized structures, it is possible (although by no means proved) that general learning mechanisms underlie all these different types of learning. It is notable that the different brain areas have many properties in common: short-term and long-term memory, one-trial and multiple-trial learning, similar cellular mechanisms and biochemical pathways, and activation of particular memory genes. This suggests the possibility of general

learning principles at the cellular and subcellular levels.

Almost all current theories of learning and memory involve some variant of the idea that efficacy of the connections between cells can be modified based on their previous activity. Such theories urge us to seek biophysical – as well as computational – descriptions of what happens at the synapse. Although the synapse has received the most experimental attention, we should begin with the caveat that the full picture of learning and memory is largely unknown (the final section of this article reviews other forms of change in the brain that could have a role in learning and memory).

The Synapse

Over a century ago, the idea that neural tissue is a continuous network, or reticulum, was challenged by the proposal that the nervous system is an intricate network of discrete cells. The great Spanish neuroscientist, Ramón y Cajal, proposed this 'neuron doctrine', which ushered in an important new idea: separate cells influence each other primarily through specialized connections called synapses. Rámon y Cajal is credited as the first to suggest that learning and memory might occur by changes in the connections between neurons.

FORMS OF SYNAPTIC PLASTICITY

What is plasticity? A plastic system is one that is changeable, and able to retain change (hence, when a manufacturer moulds a cup out of a lump of plastic, it is useful only because it retains its new shape). A nonplastic system would be unable to store memories of anything, being unchanged by its experiences.

The brain comprises both plastic and nonplastic components. The brain's control of respiration and heartbeat is not thought to be plastic, just as a bird's knowledge of how to build a nest is not something learned, but hard-wired. However, many parts of the brain are plastic, and this is what allows an animal to learn about and interact with its environment. Born in one country, you might learn to forage for food from particular broad-leafed shrubs; in another, you might learn to satisfy your cravings from a refrigerator. This learning of the proper location for the food could only be accomplished because parts of your brain were plastic. In general, plasticity is much greater in the developing animal, and decreases with age.

Associative Learning

Many classic experiments in psychology demonstrate the role of association. Every student knows about Pavlov's dogs, who salivated when they heard the bell that signaled the presentation of meat powder. The behavioral psychologist B. F. Skinner found that particular stimuli caused an organism to repeat an act more frequently. He called stimuli with this effect 'reinforcers'. Other psychologists found that by providing reinforcement in a systematic way one could shape an animal's behavior in desired directions.

What underlies these fundamental learning paradigms is the notion of 'association' – in these cases, the association of the bell and the meat, or the association between a stimulus and reward, or between a particular behavior and a punishment. An appealing idea in neuroscience has been that if an association is established between two stimuli, perhaps there is a neural substrate that should directly reflect this; but what cellular mechanisms could possibly underlie association?

In 1949, the neuroscientist Donald Hebb outlined the following hypothesis:

> When an axon of cell A is near enough to excite cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased.

In other words, if a presynaptic cell (A) consistently participates in driving a postsynaptic cell (B), the connection between them is strengthened (potentiated) (Figure 1). Hebb's hypothesis goes on to prescribe that if A consistently fails to drive B, the connection is weakened (depressed). Systems that display this behaviour are said to follow a Hebbian, or correlational, learning rule. This rule can be written as

$$\Delta w(t) = \lambda x(t) y(t) \qquad (1)$$

where, at time t, $w(t)$ represents the efficacy (weight) between two neurons, and $x(t)$ and $y(t)$ are measures of pre- and postsynaptic activity (a common measure is the rate of action potential generation, or 'firing rate'); λ is the learning rate, as it specifies how quickly the weights will react to changes in $x(t)$ and $y(t)$. The rule states that if the presynaptic and postsynaptic cells are both firing more than normal (i.e. are co-active; assume $x(t)$ and $y(t)$ represent firing rates above baseline, and hence are both positive numbers in this example), then the weight change will be positive (potentiation). If the activity $x(t)$ is high and $y(t)$ is low,

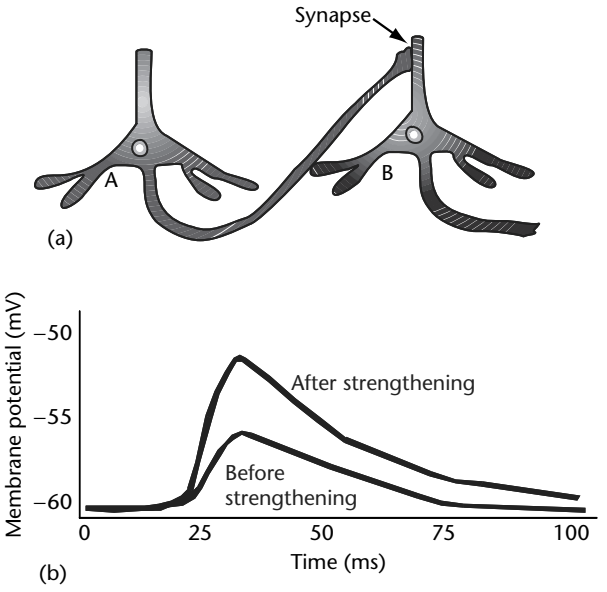

Figure 1. Learning and memory at a synapse. (a) According to Hebb's postulate, if neuron A consistently causes neuron B to fire, then the connection between them will strengthen. (b) When neuron A releases neurotransmitter on B, an excitatory postsynaptic potential (EPSP) can be measured in B. If long-term potentiation is induced by tetanic stimulation of A, the EPSP measured in B is now larger, so the connection has been strengthened.

then the connection between them will weaken. To date, most models of neural function employ such a rule.

At the time Hebb proposed his hypothesis, there was no direct experimental demonstration that could be marshaled for support. Then, in a famous experiment in 1973, researchers stimulated a bundle of nerve fibers in a rabbit's hippocampus. Trains of electrical stimulation were administered at 15 Hz for 10–15 s, and it was shown that one or two such exposures were enough to 'condition' (potentiate) an increased electrical response from the postsynaptic cell for up to 10 h. In other words, this was a direct demonstration that connections could be modified based on the history of the activity of the cells involved.

Although this experiment galvanized thousands of researchers in the ensuing decades, it is still unclear whether the results of this experiment completely and accurately represent the neural changes that take place during learning and memory. Our current understanding suggests that some of the stimuli employed in such experiments are unlikely to be seen in the real animal *in vivo*. However, other stimuli do appear to be biologically feasible; for example, it was later discovered that an optimal stimulus to induce this sort of modification is at a frequency of 5 Hz – which is the frequency of electrical rhythms seen in the hippocampus when animals explore a novel environment.

In general, the biological relevance of the correlational learning rule has been supported by decades of research into the detailed biophysical properties of synapses. It is clear from developmental neuroscience that appropriate neural activity is required for establishing the proper connections between brain areas. This developmental self-organization is consistent with the hypothesis that changes in synaptic efficacy are controlled by processes resembling Hebbian rules. However, such activity-dependent properties of cells are not limited to the developing animal: many experiments have now revealed the reorganizational plasticity of the adult brain. It is widely hypothesized that the cellular rules that account for large-scale self-organization and reorganization (in both the developing and adult brain) may be the same rules that account for learning and memory.

Different Timescales of Synaptic Plasticity

Animals display learning and memory over many different timescales; it may be significant that changes in synaptic efficacy, or strength, similarly occur over different timescales.

Short-term changes

Sometimes the modification of synaptic efficacy is short-term, lasting on the scale of seconds to minutes. The two most commonly studied examples of this are short-term potentiation and short-term depression (also known as fast synaptic depression). The former can be induced by increased levels of intracellular calcium in an axon terminal due to recent activity. This leads to a higher probability of neurotransmitter release with successive activity, and thus the connection between the cells is considered potentiated. Short-term depression can come about by a depletion of the readily releasable pool of neurotransmitter vesicles, which take time to be repackaged and docked. When high activity levels cause increased vesicle release, the terminal becomes temporarily less able to respond to future activity, and is thus considered depressed.

Such fast modifications in synaptic activity can cause rapid dynamic changes in the behavior of networks, modifying their function as a result of recent activity.

Long-term changes

Long-term potentiation

In the 1973 synaptic modification experiment mentioned above, the changes lasted for many hours, and thus the phenomenon was labelled 'long-term potentiation' (LTP). We now know that the phenomenon of LTP is common to many brain areas. It is studied most extensively in the neocortex and hippocampus, the latter because of its crucial role in memory formation.

In most cases, LTP is induced only when the activity in the postsynaptic cell (depolarization) is associated with activity in the presynaptic cell. Depolarization alone or presynaptic activity alone is ineffective. Additionally, LTP is synapse specific, which means that each individual synapse on a cell could, in principle, strengthen or weaken according to its own personal history (although this view is questioned by some).

What goes up must come down: long-term depression

If a connection is able to potentiate, it also needs the ability to depress, otherwise the system will become saturated, and be unable to store anything new. Long-term depression (LTD) is obtained by using low-frequency repetitive stimulation (e.g. 1 Hz instead of 15 Hz). After this conditioning, the connection between two cells is weakened. There are other ways to achieve LTD, as will be seen in the section on timing, below. Long-term potentiation and depression are part of the same phenomenon; LTD is found at the same synapses as LTP and also depends on the same mechanisms.

In general, it is suggested that LTP and LTD could mediate one-trial associative learning, since pairing of two events (presynaptic and postsynaptic activity) creates a long-term change in synapses. The biophysical mechanisms underlying long-term changes are still a subject of intense investigation, but the differences between LTP and LTD are mainly thought to involve differences in the concentration (and temporal dynamics) of postsynaptic calcium ions. Although it may turn out to be inadequate, a current hypothesis is that lower calcium concentrations lead to depression, whereas higher concentrations lead to potentiation.

The NMDA Receptor

For the induction of memory, one of the most important biophysical mechanisms is a subtype of glutamate receptor called the NMDA receptor (NMDA-R), so named because it is selectively stimulated by *N*-methyl-D-aspartate. How does this receptor work, and why does it enjoy such prominence in the research literature?

In most systems, the NMDA-R is crucial for induction of LTP. An animal can be taught a behavioral task, but with the infusion of NMDA-R antagonists, the ability to remember the specifics of the task disappears.

At resting potentials, the NMDA-R is blocked by magnesium. Depolarization of the postsynaptic cell expels the magnesium ions and opens up the channel. Many postsynaptic membranes contain NMDA as well as non-NMDA glutamate receptors. During normal low-frequency stimulation, only the non-NMDA channels will open, owing to Mg^{2+} blockage of the NMDA ion channels. In contrast, high-frequency presynaptic input resulting in depolarization of the postsynaptic membrane displaces the magnesium ions, making the NMDA receptors sensitive to subsequent release of glutamate. In this way, the NMDA-R can act as a coincidence detector, sensing coincidence of presynaptic and postsynaptic activity. Thus, NMDA synapses are the quintessential biological Hebbian synapses, and thus may be crucial to the storage of associations.

The fact that NMDA receptors have a particularly high permeability for calcium allows them to stimulate a second-messenger system that results in long-term structural changes to the postsynaptic cell. Interestingly, one cannot bypass the NMDA channel by depolarizing the cell to allow calcium influx. It is not simply the amount of calcium influx that matters, but also the exact spatial location: the influx must occur in close proximity to the NMDA receptors at the synapse. This highlights the specificity of computations that are performed on spatial scales smaller than we can resolve.

Traditionally, it has been thought that strong presynaptic input is sufficient for local depolarization of the membrane in which the NMDA-R sits, but we now know that when a cell generates an action potential, this potential can (under the right circumstances) propagate back into the dendritic tree. Thus, a back-propagating action potential could act as a global dendritic signal, depolarizing thousands of synapses (and their NMDA-Rs) at once.

Certain forms of LTP induction are known to depend on postsynaptic burst firing. Hippocampal burst firing is presumed to be associated with memory induction *in vivo* (it occurs during active exploration of novel environments).

Note that the NMDA-R is only necessary for the induction of most forms of LTP and LTD; other

mechanisms underlie the maintenance of the changes – most generally, new protein synthesis is required at the nucleus of the cell. An animal can be trained to associate two stimuli (say, pairing an electric shock with a bright light) in the short term, but if protein synthesis is blocked, no long-term memory develops.

How Do Synaptic Efficacy Changes Take Place?

The common observation with plasticity studies is that after conditioning, the presynaptic cell gives a stronger or weakened electrical input to the postsynaptic cell. Mechanistically, how does the strength of an individual synapse change?

A synapse, like most elements of a cell, has many modifiable parameters. Since we know the induction of most forms of conditioning to be dependent upon the NMDA-R, and since the NMDA-R is postsynaptic, this hints at postsynaptic changes. It is known that calcium ions flow in and trigger a postsynaptic biochemical cascade involving protein kinases, and that blocking specific kinases stops LTP: some kinases that have been implicated are calcium–calmodulin kinase II (CaMK-II) and protein kinase C (PKC). These cascades eventually lead to the genome, and to the synthesis of new proteins which solidify the changes – for example, by the expression of more postsynaptic neurotransmitter receptors. Recently, it has been shown that certain receptors appear to increase in concentration after LTP. Other structural changes (such as the formation of new dendritic spines) also appear to occur minutes after LTP induction.

Presynaptic changes and retrograde messengers

Changes at the presynaptic terminal could also lead to greater synaptic efficacy; for example, an increase in probability of transmitter release. There is a raging debate as to whether the changes take place pre- or postsynaptically, and as in most great debates in biology, the answer will probably turn out to be both.

Whatever the case, any presynaptic changes will of necessity require a signal passing back from the postsynaptic to the presynaptic side. Such a signal has been labeled a retrograde messenger. There are several good candidates for biologically realistic second messengers. For example, the influx of calcium (and the subsequent activation of CaMK-II) induces synthesis of nitric oxide (NO), the molecules of which are so small that they effortlessly diffuse through cell membranes to presynaptic terminals (thus acting as a retrograde messenger) to induce the presynaptic neuron to enhance transmitter release.

Very long-term storage (compression)

It is speculated that associations formed in a subset of hippocampal neurons (area CA3) are built into more economical form in another part of the hippocampus (CA1 neurons). The hippocampus itself may function by building more economical storage (long-term memory) in the cerebral cortex. In this view, the hippocampus functions as a transient intermediary in the formation of long-term memories in the association areas of the cerebral cortex. From the point of view of psychology, such long-term storage may go hand-in-hand with the observation that we improve our learning and perform tasks faster as they become proceduralized (i.e. with practice, a skill becomes faster until it is virtually automatic).

IMPORTANCE OF TEMPORAL ORDER IN SYNAPTIC PLASTICITY

Timing: More than Coincidence?

For a long time Hebb's rule was summarized as 'neurons that fire together, wire together'. In other words, if two cells fire within some small window of coincidence, the connection between them is strengthened. However, this rule turns out to be insufficient. Hebbian rules are good for forming associations, but one theoretical shortcoming is that such rules are insensitive to the order of events. Experiments have long shown that animals are strictly sensitive to the order of sensory inputs – such that, for example, Pavlov's dog will not learn an association if the meat is presented before the bell. Similarly, many animals develop strong aversion to a tasty food following a single experience of nausea after eating it, but reversing the order (nausea and then the food) does not lead to an aversion.

It is now becoming appreciated how the relative timing of presynaptic and postsynaptic activity may be crucial at the cellular level. An interesting rule has emerged about the timing of synaptic change, at least in some systems. If an input from cell A contributes to driving cell B, then the synapse is strengthened. If an input from A comes after cell B has fired, the synapse is weakened. The importance of this timing can be seen in Figure 2. This learning rule is commonly called a 'temporally asymmetric Hebbian rule', and has expanded our view of the importance of exact spike timing.

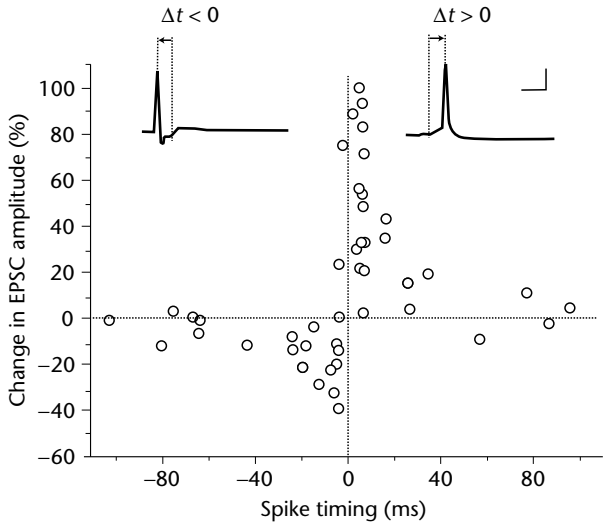

Figure 2. Timing for synaptic potentiation and depression. In this plot, reproduced from Bi and Poo (1998, *Journal of Neuroscience*), $\Delta t > 0$ means that the postsynaptic spike occurred after the presynaptic spike. In such cases the presynaptic spike has a role in the activation of the postsynaptic spike and the connection between the two neurons strengthens, as measured by the percentage change in the excitatory postsynaptic current (EPSC). If the postsynaptic spike arrives before the presynaptic spikes ($\Delta t < 0$), the connection between the two neurons weakens.

Most generally, the temporally asymmetric rule strengthens connections that are predictive; if A consistently fires before B, it can be viewed as a successful prediction, and will be strengthened. (As a caveat, it should be noted that the data in Figure 2 were obtained under specialized and perhaps biologically unrealistic circumstances. It is not yet known what would happen at a synapse with a more realistic barrage *in vivo* of pre- and postsynaptic spikes.)

THE ROLE OF CONTEXT IN LEARNING AND MEMORY

No discussion of learning and memory can take place in the absence of considerations of context – both the context into which new information can be stored, and the situational context of the animal.

Informational Context

Much of what we learn is in terms of what we already know. When you read this article, the points made here would be meaningless unless you were already equipped with an idea of how to operate a book, how to read this language, and what a brain is. Two people might look at a list of important dates in Mongolian history; if one of them already commands a richly developed cognitive model of Mongolia, the new facts are much more readily incorporated into that person's network of knowledge. At the simplest levels, this reflects associational mechanisms, as discussed above. By associating one new concept with others already known, it can be more readily stored and retrieved. Associative neural networks display this sort of property: items can prime webs of association, in the same way that the smell of coffee can immediately conjure associations with the sight of the black liquid, the sensation of warmth on the hands, the bitter taste, and so on.

Situational Context

Neural wiring can dictate which experiences result in learning and which do not. Many birds learn to form a strong emotional bonding at birth to any nearby distinctive and animate object – a process known as imprinting (this is a good example of one-trial learning). Many neural network models fail at such learning, requiring many thousands of trials for the learning of even simple tasks. More recent neural network models have therefore begun to include context as an important variable. The notion that particular types of learning are triggered to occur at certain times – 'schematic' learning – is an expression of an important philosophy that has emerged in neuroscience: the brain is not a *tabula rasa*, a blank slate upon which the world imprints itself. Instead, the brain comes preequipped for certain types of learning, and learning in particular situations. Experience is more likely to result in learning when it has relevance to the life of the organism – especially when it is connected to pleasure or pain, fear or satisfaction.

One way that context is likely to be expressed, biophysically, is through neuromodulatory systems, which are generally global neural systems that signal reward, punishment, alertness, and so on. Neuromodulatory systems have a crucial role not only in developmental plasticity, but also in the synaptic plasticity of learning and memory in the adult. Among other functions, neuromodulators can turn synaptic plasticity on and off, and influence the presynaptic/postsynaptic communication pathways. In this way the plasticity of synapses can be gated, so that learning takes place only at the appropriate time instead of each time activity passes through the cell – which could, in theory, overwrite previous learning. It has been demonstrated in the adult animal that reorganization of

parts of the cortex can occur only when paired with the release of particular neuromodulators.

Lastly, there has been a recent surge of physiological experiments in awake, behaving animals to study the role of attention. It was known by the ancient Greeks that learning and memory are most reliable when a student is paying attention. Recent experiments have shown that the firing rates of individual cells can be highly modulated by the attentional state of the animal. Since the number and timing of spikes seem to be important, it is easy to see how changes in firing rate could modify the dynamic changes in synaptic modification in the networks of cells.

OTHER FORMS OF PLASTICITY

Although synaptic transmission has been favored as the major means of communication between nerve cells, it is almost certainly an incomplete description. Signals of synaptic or nonsynaptic origin can diffuse through large volumes of neural tissue to affect signal-readers, even at distant sites. This mode of communication, 'volume signaling', can function between neurons and also between neurons and glial cells. The consideration of signaling that extends beyond the synapse allows the three-dimensional arrangement of neural elements to play a part in information processing. Such a hypothesis would predict that if one could reproduce the synaptic connections of the brain on a two-dimensional circuit board, the resulting machine would be insufficient for the functional simulation of neural tissue. This is increasingly becoming appreciated as researchers attempt to form models with explicit representations of the three-dimensional composition of neural tissue.

Although this discussion has centered on changes at chemical synapses, certain types of synapses are electrical. Electrical synapses, or 'gap junctions', are increasingly being discovered in the mammalian brain, especially within networks of inhibitory cells. In essence, they are ion channels running through one cell membrane to the cell membrane of an adjoining cell. Gap junctions permit rapid and bidirectional flow of ions between cells, and it is possible that they could exhibit plasticity or other properties that give synapses a central role in learning and memory.

There are many other possible substrates in which to store activity-dependent changes. Researchers are now studying what they call 'intrinsic' (or nonsynaptic) changes in cells: changes in the excitability of the cell, changes in the distribution of ion channels, changes in the shape of dendritic trees, changes in the phosphorylation states of intracellular proteins, and so on. With so many degrees of freedom in biological systems, the possibilities are vast for discovering different storage strategies for learning and memory.

Additionally, in many cases memory can be stored (at least short-term) in reverberating circuits of activity. For example, a form of short-term memory referred to as 'working memory' is associated with the ability to store contemporary representations of the outside world. The apparent locus for working memory is in the prefrontal area of the cerebral cortex. Monkeys shown that food is hidden behind an obscuring object, but who are restrained from immediately taking the food, maintain their knowledge of the whereabouts of the food, and will find it after a delay, once restraints have been removed. Monkeys who have had areas of their prefrontal cortex surgically removed will forget about the food as soon as they can no longer see it ('out of sight, out of mind'). The mechanisms that underlie this working memory may be in circuits of reverberating activity, which keep the information temporarily stored without the need for synaptic change.

CONCLUSION

Many features of neural plasticity in learning and memory seem consistent with Hebbian rules. This convergence of experimental and theoretical approaches provides a powerful example of the modern approach to neurobiology. Future experiments will continue to elucidate the extent to which synaptic learning rules account for the properties of learning and memory, and will contribute to expanding neural models to encompass issues of timing and context.

Further Reading

Bi GQ and Poo MM (1998) Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience* **18**: 10464–10472.

Churchland PS and Sejnowski TJ (1995) *The Computational Brain*. Cambridge, MA: MIT Press.

Koch C (1998) *The Biophysics of Computation*. New York, NY: Oxford University Press.

Montague PR and Sejnowski TJ (1994) The predictive brain: temporal coincidence and temporal order in synaptic learning mechanisms. *Learning and Memory* **1**: 1–33.