Modifiable Neuronal Connections: An Overview for Psychiatrists

Kathryn J. Jeffery, M.B., Ch.B., Ph.D., and Ian C. Reid, M.B., Ch.B., Ph.D., M.R.C.Psych.

Synaptic plasticity is currently the target of much neurobiological research, because it is thought to play an important role in brain function (particularly memory formation). However, it has attracted little attention from psychiatrists to date despite accumulating evidence that links it to various clinical syndromes, including amnesia and possibly psychosis. The purpose of this article is to present an overview of the two major arms of synaptic plasticity research—theoretical (the field of neural network modeling) and neurobiological (long-term potentiation). Artificial neural networks are a class of theoretical model that has been developed with the aim of understanding how information could, in principle, be represented by large numbers of interconnected and relatively simple units. Over the past few decades, several theoretical accounts of information-processing mechanisms have been developed, and these are briefly reviewed. The principle common to representation formation in nearly all neural networks is that of "associability"—the idea that streams of information are combined by forming, strengthening, or pruning connections between them to form new representations that can later be retrieved. Associability also lies at the heart of psychological theories of information storage in the brain. Research into associability has directed the attention of many experimenters toward the possible biological correlates of such mechanisms. Of particular interest is the recent discovery that some neurons appear to possess connections of modifiable strength. The implications of this finding for psychiatry are discussed in relation to representational disorders such as delusions and amnesia.

(Am J Psychiatry 1997; 154:156–164)

Many common symptoms encountered in psychiatric practice, such as delusions, perplexity or delusional mood, phobias, and amnesia, appear to involve abnormalities of the brain's representation of the outside world. Although explanations of these phenomena undoubtedly include some underlying disturbance of the basic machinery of perception, cognition, and knowledge representation in the brain, constructive efforts to formulate explanations in these terms have been rare (1). The hypothesis that delusions and related phenomena arise from some disturbance of the cognitive machinery of belief, memory, and their associated neural representations provides a pointer to those areas of inquiry that could shed light on their biological underpinnings. The theoretical and experimental study of neural representation has mushroomed in recent years; in this article we attempt to take account of those developments that may prove to be relevant to psychiatry.

Received March 1, 1996; revision received Aug. 14, 1996; accepted Aug. 16, 1996. From the Department of Anatomy and Developmental Biology, University College London. Address reprint requests to Dr. Jeffery, Department of Anatomy and Developmental Biology, University College London, Gower Street, London WC1E 6BT, U.K.; kate@anatomy.ucl.ac.uk (e-mail).

The authors thank Neil Burgess, Geoff Goodhill, and David Willshaw for their comments.

Although psychiatry has tended to leave aside the mechanistic details of representation formation in the brain, these details have attracted attention from such disparate fields as neuroscience, psychology, computing, and, perhaps surprisingly, physics and engineering. These disciplines have in recent decades devoted much attention to the question of what happens to information once it has been transduced by the sensory receptors and passed into the brain. There are several reasons for this abundance of interest. The brain is an enormously complex structure, and it has only been with the comparatively recent development and evolution of computers that it has been possible to muster the computational power needed to simulate how large numbers of small neuron-like processing elements could represent "knowledge." In addition, research efforts have been motivated by the inability (to date) of artificial intelligence to replicate convincingly some useful, and apparently trivial, brain functions such as face recognition and speech synthesis—functions that our own brains perform with ease. The failure of classical engineering and computer approaches to solve these apparently simple problems has forced scientists to look more carefully at how real brains do this. One result has been the development of so-called artificial neural networks, which are constructed according to biologically in-

spired principles, and whose properties are investigated in the hope not only that they will provide solutions to recalcitrant engineering and computing problems but also that they will shed light on brain function.

The modern field of neural network research was greatly influenced by the insight of the pioneering neuroanatomist Cajal that information could be stored by modifying the connections between communicating nerve cells, in order to form associations (2). This idea was formalized by Hebb (3), who suggested that such modifications should take place between the connected cells if (and only if) both neurons were simultaneously active. Following the development of the earliest neural networks, many of which were inspired by Hebb's theoretical postulate, physiologists discovered that some memory-associated structures possessed modifiable synapses. It appeared that the brain therefore had at least some of the properties needed to implement Hebb's information storage paradigm. Subsequently, both neural network research and memory research have proceeded in parallel to elucidate the theoretical properties of ideal neural networks and the actual properties of information storage in the brain. It seems increasingly likely that synaptic plasticity does indeed play a critical role in representation formation, and it is therefore probable that phenomena such as delusions and amnesia, which are essentially disorders of representation, will eventually be found to involve the mechanisms of neural plasticity in one form or another.

First, we review the field of neural network research and describe how artificial neural networks are constructed from large numbers of simple neuron-like elements, how they process and store information, and how their brain-like properties arise. Second, we explore some of the recent discoveries concerning representation in real brains, including the molecular biology that endows synapses with their properties of Hebbian modification. These discoveries are considered to be important because they constrain the types of neural network models that can be said to possess biological relevance and because they provide a potential target for therapeutic intervention. Intervention at sites of synaptic plasticity is already promising to be useful in many neurological conditions, and an understanding of the molecular biology of representation formation may provide therapies that are also useful to psychiatrists. Finally, we review recent evidence that a disturbance of synaptic plasticity may underlie a well-established psychiatric condition, postelectroconvulsive amnesia.

THEORETICAL ASPECTS OF SYNAPTIC PLASTICITY: ARTIFICIAL NEURAL NETWORKS

The Hebb Rule

An artificial neural network is a model of how a group of neuron-like elements might behave when connected together in various ways and made to influence each other according to certain rules. The aim is to simulate brain-like behavior in a simplified manner that is then amenable to analysis.

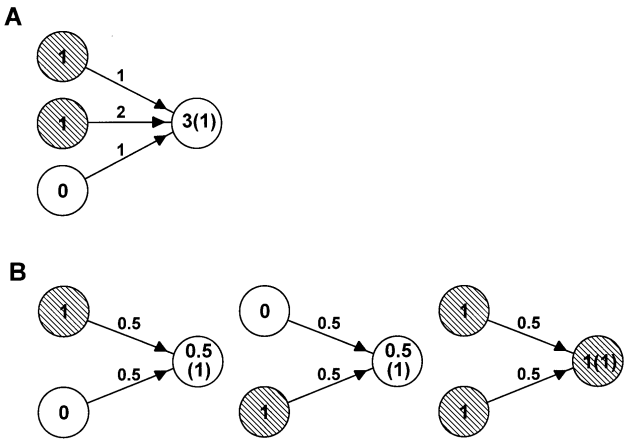
Early in this century Cajal proposed the "neuron doctrine" (2), part of which (the "principle of dynamic polarization") states that impulse conduction from one neuron to another proceeds in one direction only. This property of one-way transmission between neurons forms the basis of nearly all artificial neural networks. Cajal also proposed the novel idea that information could be stored by modifying interneuronal connections. This principle was formalized by Hebb, who proposed that the connection from one neuron to another, rather than being a fixed, passive conductor like a piece of wire, could be increased in strength when both neurons were simultaneously active, so that a neuron could subsequently be made to excite the next one more easily than before (3). Specifically, he put forward his now well-known Hebb rule: "When an axon of cell A is near enough to excite a cell B and repeatedly and persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficacy, as one of the cells firing B, is increased" (3, p. 62)

It may be seen that the Hebb rule allows for the association of neural events. If cell A is connected to cell B, and they are both active simultaneously, and the connection between them is therefore strengthened, then the future occurrence of activity in cell A is more likely to produce activity in cell B as well. For example, if cell B is one of a group of cells whose activity is signaling food, and cell A is part of the representation of the sound of a bell ringing, then if the connection between them is strengthened in obedience to the Hebb rule, the next time the bell is rung, the food representation will be more easily activated. If this connection is sufficiently strong, the food representation may be evoked by the sound of the bell even in the absence of food. This, of course, is the familiar paradigm of Pavlov. This simplified example illustrates how the Hebb rule imparts the property of association of activity to groups of neurons. Thus, as well as one-way conduction of impulses, a second property common to nearly all neural networks is modifiability of the connections (also called synaptic weights) between neurons (or units) according to rules.

Properties of Neural Networks

Information storage and retrieval by the brain have some curious properties. First, for some kinds of problems they are very fast: for example, if a subject is shown a photograph of a face and asked whether he or she has ever seen that person before, a reply will probably be forthcoming in less than a second. Currently, modern computers are unable to match this performance using processing elements many thousands of times faster than the brain's comparatively sluggish synapses. Second, even serious brain damage usually appears not to remove specific distant memories. These two features exemplify the differences between information stored by the brain and

FIGURE 1. The McCulloch-Pitts Model[a]



[a]A) Three input neurons connect to an output neuron through synapses of fixed strength. The numbers outside parentheses indicate neuronal activation or synaptic strength, and the number inside parentheses indicates the firing threshold of the output neuron. Firing neurons are shaded. Activation of the output neuron is determined by multiplying the activation of each active input by the strength of the connection between them and then adding the totals. Provided its activation equals or exceeds the threshold, the neuron will fire. B) Adjustment of the postsynaptic threshold allows a configuration of neurons to perform an "AND" operation. Firing of only the top neuron (left) or only the bottom neuron (middle) fails to activate the output neuron past its firing threshold. However, firing of both input neurons together (right) produces a postsynaptic activation strong enough to induce the cell to fire.

that stored by a computer and are called, respectively, content-addressability and distributed representation. Content-addressability means that retrieval of information does not require an exhaustive search through the memory store (which, in the example given, would require many hours in order to check every face ever seen) but that features of the item to be retrieved are used to activate only those items in memory that possess the same features (or a subset of them—for example, only young women with long brown hair). Distributed representation refers to the observation that a given memory is probably stored across many brain regions. Evidence for this includes the finding, first reported by Lashley (4), that experimental removal of even large areas of the cerebral cortex does not appear to result in the selective obliteration of memories, although patients with extensive brain damage may report that memories in general become less clear. The gradual (rather than abrupt) decline of the functioning of a system with increasing injury is often referred to as graceful degradation. In the brain, it is a byproduct of the distributed nature of its information storage.

Artificial neural networks possessing the two properties of one-way impulse traffic and modifiable connections (according to the Hebb rule or variants of it) have turned out to have some intriguingly brain-like properties, including content addressability, distributed representation, and graceful degradation. This has made them of great interest in the development of a theoretical understanding of brain function.

Early Artificial Neural Networks

The first attempt to produce "behavior" from artificial neuron-like elements linked together was made by McCulloch and Pitts more than 50 years ago (5). At this time the first computers, using the binary logic that is still used today, were being developed, and McCulloch (a neurophysiologist) and Pitts (a logician) were struck by the way in which neurons themselves appeared to behave in a binary fashion—active or silent. They suggested that the on-or-off property of neurons could, as in a computer, be used to perform logical operations. The McCulloch-Pitts model (figure 1, part A) therefore consisted of a group of neurons connected to a processing neuron whose task was to add up the strength of all of its inputs and determine whether the total was enough to exceed its threshold for firing an action potential. By using such a network, it is possible to perform a number of logical operations. Figure 1, part B, illustrates how a McCulloch-Pitts neuron performs an "AND" operation. In this type of computation, the desired response of the output neuron is that it should fire if both of its input neurons are active together, but not if only one or the other is active alone. This response can be achieved by setting the threshold of the output neuron such that neither input is strong enough to activate the output neuron by itself. Simple neural network models are based on the McCulloch-Pitts neuron, with the added feature that their connections can change strength, as Hebb suggested, so that a judicious combination of connection strengths and thresholds may allow an output neuron to fire correctly. Figure 1 illustrates how the AND operation can be represented with the use of variable connection strengths.

There are two serious limitations to this type of model. The first is that although it is possible for an outside observer to see what connection strengths might enable a given input to result in a desired output, the network cannot learn these connections for itself—they have to be set by hand, by someone who knows in advance the nature of the problem to be solved. This kind of "hardwiring" may satisfactorily explain those types of nervous system connectivity that are designed to solve a constant problem, such as edge detection in the retina, but it does not explain how a network might learn, for example, the Pavlovian association paradigm. The second difficulty is that with only an input and an output layer of neurons, as in the McCulloch-Pitts model, there are some problems, such as the "exclusive OR," that simply cannot be solved by any straightforward combination of inputs and connection strengths. The exclusive OR is analogous to teaching Pavlov's dog to expect food after either a bell or a light but not after both a bell *and* a light. It requires that a neuron learn to fire when either one of its two inputs is active but not both of them. It can be seen in figure 2 that this would require a self-contradictory wiring arrangement. The technical term for this type of problem is linear inseparability.

The task facing neural network researchers, then, is how to build a neural network that is able to find for

itself a suitable wiring arrangement to perform a desired computation. Of course, in principle, finding such a solution does not necessarily mean that the brain uses the same mechanisms to process information. Neural network researchers with an interest in biological plausibility must therefore look to the results of neurophysiological studies to guide the development of network models. This is discussed further in the section on Biological Synaptic Plasticity.

Representation Formation by Error Correction

One of the first neural networks that could "learn" was built by Rosenblatt (6) in 1962 and called the perceptron. In order to build a network that can set its own synaptic weights without the need for outside intervention, it is necessary to build into it some kind of rule about how and when to change connections. The perceptron learning rule owes its origins to the Hebb rule and is of a type known as error correction. In effect, the output neuron is given feedback about whether it has fired correctly or not. If it has fired when it should have remained silent, then those inputs that have been active at the time are treated as having told it to fire inappropriately, and their connections are therefore decreased in strength. If the output neuron has remained silent when it should have fired, then the connections coming from the active inputs are strengthened. If this cycle of input, output, and weight change is repeated enough times, it can be shown that the connections will eventually arrange themselves into a suitable configuration for solving the problem being presented, provided it is not linearly inseparable (7).

Associative Memory

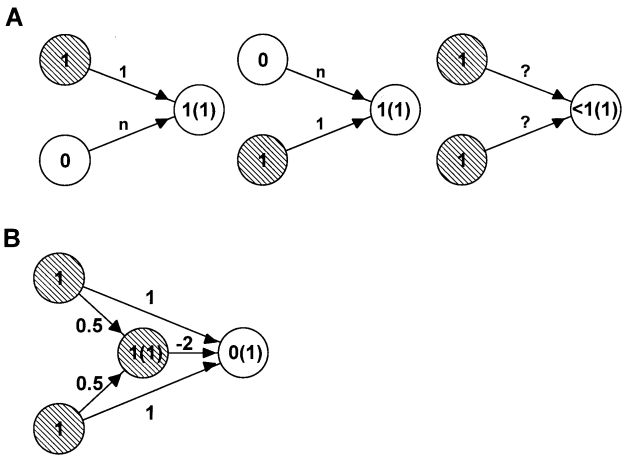
In 1969 Willshaw and colleagues (8) showed that a neural network with many output units, instead of only one, could function as an associative memory. In other words, it could take repeated presentations of a pattern of input activity and, using a Hebb-type rule, adjust its connections according to a desired pattern of output activity so that, eventually, presentation of the input pattern would produce the second pattern as the output. The associative network has excited much speculation that some brain regions might use a similar principle to perform associations such as the Pavlovian example we have given. However, the great difficulty with the associative network is that it, too, is unable to solve linearly inseparable problems.

The problem of linear inseparability has been tackled by modifying network architecture—that is, the particular arrangement of neurons over which the learning rule is operating. In the case of the exclusive OR, the paradoxical wiring problem can be solved by introducing another neuron, whose task is to resolve the conflict.

The layer of neurons that is interposed between the input and output neurons, in order to untangle paradoxes and convert a linearly inseparable problem into a set of smaller linearly separable ones, is called a hidden layer. With enough neurons in a hidden layer, almost any problem can be solved. One difficulty lies in knowing how many hidden-layer neurons is enough. A more serious problem is how to train (i.e., find the right configuration of connection strengths for) the hidden-layer connections. The problem is that an error-correcting rule will only operate when a neuron knows what it is "supposed" to be doing—in other words, is receiving some form of feedback about whether or not it has fired appropriately. A hidden-layer neuron, however, does not know what it is supposed to be doing, because its task is to do whatever is necessary to make the output neuron fire correctly. If it knew a priori what to do, then there would be no need to have it in the network. An ingenious way around this problem is the technique of back-propagation of errors, developed by Rumelhart and associates (9). In this kind of network, the error (the fact that the output neurons are firing incorrectly) is passed back to the hidden layers, which also adjust their weights according to an error-minimizing rule and then pass the error back to the preceding layer, and so on.

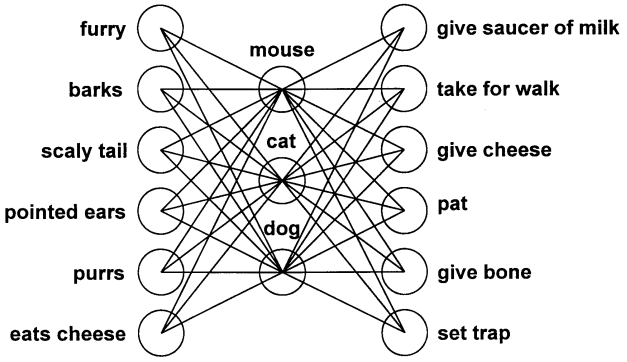
Back-propagation is quite successful at solving certain kinds of problems and is now widely used in the design of adaptive (learning) systems. However, it worries many biologists, who feel that there is no evidence that this type of error propagation occurs in real brains. Nevertheless, analysis of the results of back-propagation has revealed that when the network has settled down to a stable state and no more changes are taking place (that is, the connections to all of the neurons have arranged themselves into a suitable configuration), it

FIGURE 2. The "Exclusive OR" Problem[a]



[a]A) A simple neural network cannot solve the exclusive OR problem. The output neuron can be induced to fire when either the top (left) or bottom (middle) input neuron is active, provided the synaptic strength equals or exceeds 1. However, if the conditions for firing to either input alone are satisfied, then the conditions for *not* firing to both of them cannot be met (right). n=an unspecified value. B) A solution to the exclusive OR problem is the introduction of an inhibitory hidden unit. The hidden unit will only be activated by simultaneous firing of both of the inputs, and its action is to inhibit postsynaptic excitation produced by the two input neurons.

FIGURE 3. Structure of a Back-Propagation Network With Three Hidden Units[a]



[a]After training, the connections become weighted in such a way that each hidden unit comes to represent the cluster of features associated with one of the objects presented to the network as input, and it will in turn excite the appropriate target responses.

can be seen that the hidden-layer neurons have come to form an internal representation of the problem, using groupings of the data that were not present in the information originally presented to the network (figure 3).

This type of "concept" formation is an emergent property of the network and could not have necessarily been predicted from observation of the input pattern. Thus, although the brain probably uses somewhat different algorithms, analysis of artificial neural networks has shown that it is in principle possible to produce complex-looking behavior from very simple units acting together in large numbers.

Representation by Stable States

Back-propagation and other kinds of error-correcting algorithms have provided ways in which a neural network might learn to associate patterns of activity, forming internal representations of the problem by using hidden units. In this type of learning, the output units are being told what to do by an external agency. A real-world example may be found in some neural network models of the cerebellum, where it is postulated that motor commands are processed with the use of feedback from muscles and nerves as a "teacher" to decide whether the command has been executed correctly (according to intention) or needs adjustment (10).

In some neural network models, the final configurations of weights and responses of the output neurons are discovered by the network itself and not forced upon it by means of any predetermined output requirements. One of the most interesting examples of this type of network was introduced by Hopfield (11). Hopfield was intrigued by the way in which large systems in nature often appear to generate spontaneously organized behavior, even when they are composed of many interacting elements. An example of such a system is the Ising "spin glass" model for dilute magnetic alloys.

An Ising spin glass is composed of a large number of particles, some of which carry a small magnetic field and are therefore able to influence the alignment of the other magnetic particles within the material. Whether a particle will influence surrounding particles to orient themselves in the same or the opposite direction depends on the distance between them; thus, each particle may receive conflicting influences from its neighbors. The amount of conflict is referred to as "frustration," and the material as a whole, when left to its own devices, will eventually settle into a stable state in which the total amount of frustration is minimized. This property arises from the cooperative interaction of a large number of particles able to exert effects on each other.

In an Ising spin glass, the driving force for the development of stable states is the tendency of the material to minimize its total energy. Hopfield realized that a large number of interacting neurons could also be thought of as having an "energy" and therefore be subject to the same tendency to find stable states in which the energy was lowest. These low-energy configurations could in principle serve as memories. If the system started near a stable state, then over time it would naturally gravitate to that state, following the path of least resistance. An intuitive way to see this is to imagine the possible states of the network as forming a hillocky terrain, with the hills representing energy levels and the current state of the system being represented by a ball rolling down one of the hills. Retrieval of a memory is analogous to placing the ball near a hollow and following it until it stops rolling. Such gradient-descent methods of retrieval effectively serve as content-addressable memories, because presentation of part of the memory to be retrieved will place the system in a state somewhere near a hollow (a local minimum) and allow the full memory to be retrieved by energy minimization.

Hopfield's model is biologically implausible, not least because to allow stable states to evolve reliably, it requires that the neurons be symmetrically interconnected so that each neuron both sends and receives weights of equal strength—something that is not observed in the brain. A more serious problem is that such systems tend to find false minima: that is, the ball rolls into a nearby hollow that is not the "right" one. This problem worsens as the number of hollows increases (i.e., as more and more memories are stored), and eventually the network becomes so overloaded with memories that even previously stored ones become irretrievable, unless some mechanism for forgetting is introduced. Nevertheless, the Hopfield model has introduced an important new category of neuronal representation.

Competitive Learning

Another type of spontaneous representation formation by unsupervised learning is competitive learning (12–14). A competitive network takes a series of inputs and forms categories in such a way that similar inputs are placed in the same category and dissimilar inputs in different categories. In real-world terms, for example, this might amount to learning, after being presented

with inputs of a sequence of apples and lemons, that although each individual apple will have a subtly different taste, color, and texture from that of other previously encountered apples, it will be more similar to other apples than to the lemons. In neuronal terms this means that because each new apple encountered evokes a pattern of activity across the input units that is more closely related to the activity evoked by other apples than to that evoked by any of the lemons, the responses of the output neurons can be organized into groups corresponding to the real-world categories.

A network that can do this kind of classification is able to generalize to examples it has never seen before, such as a completely new apple. Furthermore, it can take ambiguous examples (such as a green, crunchy, and sour-tasting fruit) and make a sensible decision about which category to place it in—something that traditional computer algorithms often find hard to do. This kind of behavior therefore more closely approximates brain-like processing than conventional computer-like processing.

BIOLOGICAL SYNAPTIC PLASTICITY

Although neural network research has always focused attention on the structure of network representations, biologists have only recently begun to examine how knowledge is represented by living brains. The general anatomical locus of thoughts, concepts, memories, and emotions remains unknown, although the development of functional brain imaging techniques is gradually localizing some types of representation to various cortical regions. More important from the point of view of synaptic plasticity research, the microstructure of representations also remains unknown; it has yet to be established, for example, whether memories are stored in something resembling an associative or competitive network, in a stable-state configuration like a Hopfield network, or in an entirely novel manner. Recently, however, both theoreticians and biologists have begun to try to bridge the divide between artificial and natural neural networks.

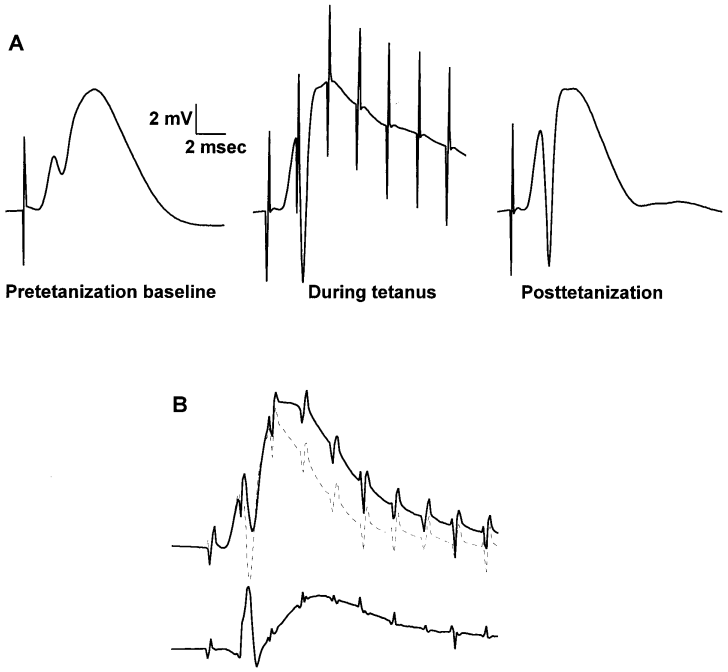
Prominent among theoretical approaches to the understanding of psychiatric conditions have been recent efforts to model the symptoms of schizophrenia by introducing "pathology" into artificial neural networks. Cohen and Servan-Schreiber (15) used back-propagation networks to model performance on three psychological tests of contextual processing on which the performance of schizophrenic subjects is known to be impaired. They introduced a change in the networks analogous to reducing dopamine in the prefrontal cortex (the brain region thought to mediate performance on these tasks) and found that the networks' performance degraded in a way very similar to that of the schizophrenic subjects. They suggested that their models may also provide "a framework for exploring the role of neuromodulatory systems in other forms of illness" (p. 68). Ruppin et al. (16) used an attractor network to model positive psychotic symptoms in persons with schizophrenia. In their model, synaptic degeneration in the input pathway was accompanied by an increase in local connections (corresponding to reactive synaptogenesis). Memory retrieval under these circumstances was relatively preserved, but spontaneous activation of noncued memory patterns occurred when either the internal synaptic strength or noise increased beyond a certain level. This biased spontaneous retrieval (analogous to delusions and hallucinations) tended to be self-limiting as a global attractor state formed, mimicking the progression from positive to negative symptoms in schizophrenia. Spontaneous retrieval was also self-reinforcing, as are untreated psychotic symptoms in young persons with schizophrenia. The model generated several predictions about schizophrenia that could be tested experimentally, including synaptic compensation, increased spontaneous neural activity, and environmental cuing of delusions and hallucinations. It is hoped that a theoretical understanding of schizophrenic processes may point the way toward better-targeted treatments.

A great deal of the experimental work linking modifiable synapses to macroscopic brain behavior has been carried out in the hippocampal formation, a part of the limbic system that has long been implicated in memory formation and that more recently has been shown to be involved with the formation (and possibly storage) of a representation of the spatial environment (the so-called cognitive map [17]). There are several reasons for devoting attention to the hippocampus. First, since space is a relatively primitive concept, an understanding of how it is represented by the brain may shed light on how more complex and abstract representations, peculiar to humans, are stored. Second, the hippocampus is the principal site of degeneration in Alzheimer's disease, in which amnesia is a prominent and early symptom. An understanding of the physiology of the hippocampus may open up therapeutic possibilities in this area. Third, the strength of its synapses has been found to be readily modifiable, suggesting that here may be a good place to search for mechanisms of memory storage. Below, we present some of what is known about the molecular mechanisms of hippocampal synaptic plasticity, before concluding with recent findings which suggest that a disturbance of such plasticity may underlie some psychiatric conditions.

Synaptic Modifiability in the Hippocampus

Although Hebb proposed his synaptic modification rule in 1949, it was not until nearly 20 years later that neurobiologists discarded their model of neurons as McCulloch-Pitts elements with fixed synapses and began to look for physiological evidence of variability in synaptic strength in real neurons. In 1973 a group of researchers published the first detailed reports of artificially induced modification of synaptic strength in the hippocampus (18, 19). They focused attention on the hippocampus because of its suspected role in memory

FIGURE 4. Induction of Long-Term Potentiation[a]



[a]A) Change in synaptic strength in a population of hippocampal neurons following Hebbian convergence of activity. Left: a single electrical stimulus to the perforant path input fiber bundle produces a characteristic evoked response because of the synchronous activation of a large population of postsynaptic granule cells in the dentate gyrus. Middle: application of a high-frequency train of pulses causes massive postsynaptic depolarization, which, when combined with the presynaptic transmitter release, provides the conditions under which long-term potentiation will be induced. Right: after a high-frequency train, a single electrical stimulus of the same intensity as before now evokes a larger response, reflecting the synaptic strengthening that underlies long-term potentiation. B) The mechanism by which high-frequency stimulation triggers induction of long-term potentiation is provided by the NMDA receptor. Top: the solid line illustrates the normal response evoked during a high-frequency pulse train. The dotted line shows the response when NMDA receptors have been blocked by intraperitoneal administration of the NMDA receptor blocker 3-(2-carboxypiperazin-4-yl)-1-propenyl-1-phosphonic acid (CPP-ene). The bottom trace shows the difference between these responses, which reflects the contribution of NMDA receptor current to the evoked response during high-frequency stimulation.

formation, and they found that if the perforant path was stimulated with strong high-frequency electrical pulses, then the synapses of those fibers onto the hippocampal cells became measurably stronger and stayed so for many weeks (18, 19). Figure 4 illustrates the increase in evoked response following high-frequency stimulation of the perforant path. This phenomenon was called long-term potentiation, and subsequent studies showed that long-term potentiation could only be induced, by analogy with the Hebb rule, if the input fibers were stimulated while the postsynaptic cells were already active (20, 21). The discovery of Hebb-like synaptic plasticity in a putative memory structure was arguably one of the most important neurophysiological finds of the decade. There is now a great deal of experimental (albeit not incontrovertible) evidence linking hippocampal long-term potentiation to memory formation (22–27), and for this reason long-term potentiation is currently regarded as the best existing memory model.

Molecular Biology of Long-Term Potentiation

The molecular biology of long-term potentiation is of interest because it may possibly explain how the computational properties of plastic synapses arise—information that is needed by neural network modelers if they are to construct biologically realistic networks. A detailed review of the status of research into the mechanisms underlying long-term potentiation has been presented by Bliss and Collingridge (28). Here we restrict discussion to an explanation of how hippocampal neurons detect the association of events needed to trigger synaptic strengthening.

A considerable body of evidence now demonstrates that the necessary conditions for induction of long-term potentiation are the following: 1) the postsynaptic cell must be active (i.e., sufficiently depolarized, though not necessarily firing), and 2) the presynaptic axon terminal must have released neurotransmitter (L-glutamate). If enough presynaptic axons have been active simultaneously (a condition known as the cooperativity requirement [29]), then if their postsynaptic target is depolarized (the associativity requirement), their synapses will be strengthened. For the purposes of inducing long-term potentiation, it does not matter how these two conditions come about; for example, the postsynaptic cell may have been artificially depolarized by current injection (21), or presynaptic activity may have been mimicked by iontophoretic glutamate application. A further property of long-term potentiation, input specificity, is that other axons terminating on the same cell will not themselves develop long-term potentiation unless they too released neurotransmitter shortly before the postsynaptic cell became active (20, 30).

The signal for long-term potentiation to occur is a sudden influx of calcium ions into the postsynaptic cell; if this is blocked—for example, by intracellular injection of a Ca^{2+} chelator—then even if the associativity and cooperativity requirements are met, induction of long-term potentiation will not occur. The receptor responsible for regulating Ca^{2+} influx in response to the conjunction of events is the *N*-methyl-D-aspartate (NMDA) subtype of excitatory amino acid receptor. The NMDA receptor is unusual in that it will not open its ion channel in response to its ligand until the cell upon which it resides has been depolarized. This is because under resting membrane conditions, the channel is normally blocked by Mg^{2+} ions (31). When the cell is depolarized, these ions vacate the channel and calcium is free

to flow into the cell. Because of the requirement for strong postsynaptic depolarization to occur nearly simultaneously with presynaptic transmitter release, long-term potentiation will only occur under unusual conditions, such as those produced by a high-frequency tetanus (figure 4). It is this simple mechanism that gives long-term potentiation its Hebbian properties and, in the behaving animal, according to current theories, enables memory formation to occur.

Synaptic Plasticity, "Associability," and Psychiatric Disorder

Investigation of the phenomenon of long-term potentiation has led to the hypothesis that under normal circumstances, the NMDA receptor may participate in the formation of brain representations, perhaps through a mechanism of changes in synaptic strength analogous to those underlying long-term potentiation. It follows that abnormal function of these receptors might contribute to pathological disorders of representation. In support of this hypothesis, abnormal "associability" is a characteristic feature of many kinds of psychiatric symptoms. For example, delusions are characterized by a striking inability of the patient to be reasoned out of an aberrant belief, plus a gradual spreading of the belief system to incorporate new information as the disease progresses. It is as if the associations between ideas that form the framework of a normal belief system have become so immutable that the usual processes of assimilation and incorporation of new information can no longer operate. This observation leads to the hypothesis that mechanisms of synaptic plasticity may be abnormal in these patients. In this light, it is interesting that addition of a Hebbian component to the synaptic responses in the schizophrenia attractor model of Ruppin et al. (16) produced a retrieval bias toward just a small number of stored patterns—perhaps the network equivalent of a delusional system! Clearly, many insights into mental processes may be generated by such attempts to model them artificially.

Post-ECT Amnesia and Long-Term Potentiation

Unilateral ECT is commonly used in the treatment of severe unipolar depression, and one of the most common side effects is a marked (though transient) disorder of memory function. Since memory storage is putatively mediated by synaptic plasticity, we investigated the possibility that electroconvulsive stimulation might be associated with either a change in synaptic strength or a change in the capacity of these synapses to support long-term potentiation. Such a finding would support the hypothesis that changes in synaptic connections underlie the formation of brain representations and also suggest some therapeutic possibilities for amnesia after ECT in depressed patients.

Rats given a 10-day course of electroconvulsive stimulation similar to that used in psychiatric practice showed a large rise in the size of the evoked dentate potential that strongly resembles long-term potentiation (32, 33). This increase had a prolonged time course, comparable to that of ECT-induced amnesia (33, 34), and subsequent induction of long-term potentiation in the same pathway was impaired. Because it is known that repeated induction of long-term potentiation eventually raises synaptic strengths to a maximum level, beyond which they cannot be increased experimentally, this occlusion of long-term potentiation suggests that the changes after electroconvulsive stimulation reflected true induction of long-term potentiation, and that the failure of subsequent tetanization to induce further long-term potentiation was due to saturation of synaptic strengths. This raises the possibility that the amnesia seen after repeated ECT arises because synaptic strengths are pushed to their ceiling levels, thus preventing the formation of new memories until such time as the strengths decay back to a baseline level.

Because induction of long-term potentiation depends on NMDA receptors, an NMDA antagonist might also be expected to block the post-ECT changes in evoked potentials. As predicted, pretreatment with ketamine (a potent NMDA antagonist) prevented the electroconvulsive stimulation-associated increase in evoked response size (35). We are currently investigating the possibility that amnesia after ECT in humans might be ameliorated by administration of ketamine anesthesia before treatment. Besides its benefits for patients undergoing ECT, such a finding would constitute important support for the hypothesis that changes in synaptic strength mediate memory storage.

CONCLUSIONS

This article provides an overview of current work on representation formation by both artificial and biological neural networks. The issue of how the brain represents information at a neuronal level is one of the most important questions in neurobiology at the present time. Psychiatry has tended to divide its attention between levels well below the neuronal, such as the molecular biology of neurotransmitter release, or well above the neuronal, such as the formulation of psychoanalytic models. The middle ground has traditionally been turned over to psychology. However, recent discoveries regarding the mechanisms of certain types of cognitive processes are proving to be relevant to some psychiatric disorders.

Because of the large size (in information terms) of the brain and the current paucity of theories regarding the functioning of complex dynamic systems, an understanding of how groups of neurons can form representations cannot come about by observation or even experiment alone, but requires that experiments interact with theory to produce testable predictions on both sides. This requirement has been one of the strong motivating forces behind the rapid development of the field of neural networks, which has brought together several disciplines to design and study complex models exhib-

iting brain-like behavior. In turn, the findings of these models are motivating experimental research into biological representation formation, an important area that we hope will prove to be fertile ground for the discovery and development of new therapies for psychiatric disorders.

REFERENCES

1. McKenna PJ: Memory, knowledge and delusions. Br J Psychiatry 1991; 159:36–41
2. Cajal SR: Histologie du Système Nerveux de l'Homme et des Vertébrés, Vol 2 (1911). Translated by Azoulay L. Madrid, Instituto Ramon y Cajal, 1952
3. Hebb DO: The Organization of Behavior. New York, John Wiley & Sons, 1949
4. Lashley KS: In search of the engram, in Society of Experimental Biology Symposium Number 4: Psychological Mechanisms in Animal Behavior. London, Cambridge University Press, 1950
5. McCulloch WS, Pitts W: A logical calculus of the ideas immanent in nervous activity. Bull Mathematical Biophysics 1943; 5:115–133
6. Rosenblatt F: Principles of Neurodynamics. New York, Spartan, 1962
7. Minsky ML, Papert SA: Perceptrons. Cambridge, Mass, MIT Press, 1969
8. Willshaw DJ, Buneman OP, Longuet-Higgins HC: Non-holographic associative memory. Nature 1969; 222:960–962
9. Rumelhart DE, Hinton GE, Williams RJ: Learning representations by back-propagating errors. Nature 1986; 323:533–536
10. Marr D: A theory of cerebellar cortex. J Physiol 1969; 202:437–470
11. Hopfield JJ: Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 1982; 79:2554–2558
12. von der Malsberg C: Self-organizing of orientation sensitive cells in the striate cortex. Kybernetik 1973; 14:85–100
13. Grossberg S: Adaptive pattern classification and universal recoding, part I: parallel development and coding of neural feature detectors. Biol Cybernetics 1976; 23:121–134
14. Kohonen T: Self-Organization and Associative Memory. Berlin, Springer-Verlag, 1984
15. Cohen JD, Servan-Schreiber D: Context, cortex and dopamine: a connectionist approach to behavior and biology in schizophrenia. Psychol Rev 1992; 99:45–77
16. Ruppin E, Reggia JA, Horn D: A neural model of delusions and hallucinations in schizophrenia, in Advances in Neural Information Processing Systems, vol 7. Edited by Tesauro G, Touretzky D, Leen T. Cambridge, Mass, MIT Press, 1995, pp 149–156
17. O'Keefe J, Nadel L: The Hippocampus as a Cognitive Map. London, Oxford University Press, 1979
18. Bliss TVP, Gardner-Medwin AR: Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetised rabbit following stimulation of the perforant path. J Physiol 1973; 232:357–374
19. Bliss TVP, Lømo T: Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetised rabbit following stimulation of the perforant path. J Physiol 1973; 232:331–356
20. Andersen P, Sundberg SH, Sveen O, Wigström H: Specific long-lasting potentiation of synaptic transmission in hippocampal slices. Nature 1977; 266:736–737
21. Wigström H, Gustaffson B, Huang Y-Y, Abraham WC: Hippocampal long-term potentiation is induced by pairing single afferent volleys with intracellularly injected depolarizing current pulses. Acta Physiol Scand 1986; 126:317–319
22. Barnes CA: Memory deficits associated with senescence: a neurophysiological and behavioral study in the rat. J Comp Physiol Psychol 1979; 93:74–104
23. Morris RGM, Andersen E, Lynch G, Baudry M: Selective impairment of learning and blockade of long-term potentiation by an N-methyl-D-aspartate receptor antagonist, AP5. Nature 1986; 319:774–776
24. Laroche S, Doyère V, Bloch V: Linear relation between the magnitude of long-term potentiation in the dentate gyrus and associative learning in the rat: a demonstration using commissural inhibition and local infusion of an N-methyl-D-aspartate receptor antagonist. Neuroscience 1989; 28:375–386
25. Deupree D, Turner D, Watters C: Spatial performance correlates with in vitro potentiation in young and aged Fischer 344 rats. Brain Res 1991; 554:1–9
26. Jeffery KJ, Morris RGM: Cumulative long-term potentiation in the rat dentate gyrus correlates with, but does not modify, performance in the water maze. Hippocampus 1993; 3:133–140
27. Maren S, Patel K, Thompson RF, Mitchell D: Individual differences in emergence neophobia predict magnitude of perforant-path long-term potentiation (LTP) and plasma corticosterone levels in rats. Psychobiology 1993; 21:2–10
28. Bliss TVP, Collingridge GL: A synaptic model of memory: long-term potentiation in the hippocampus. Nature 1993; 361:31–39
29. McNaughton BL, Douglas RM, Goddard GV: Synaptic enhancement in fascia dentata: cooperativity among coactive afferents. Brain Res 1978; 157:277–293
30. McNaughton BL, Barnes CA: Physiological identification and analysis of dentate granule cell responses to stimulation of the medial and lateral perforant pathways in the rat. J Comp Neurol 1977; 175:439–454
31. Nowak L, Bregestovski P, Ascher P, Herbet A, Prochiantz A: Magnesium gates glutamate-activated channels in mouse central neurones. Nature 1984; 307:462–465
32. Stewart CA, Reid IC: Electroconvulsive stimulation and synaptic plasticity in the rat. Brain Res 1993; 620:139–141
33. Stewart CA, Jeffery KJ, Reid IC: LTP-like synaptic efficacy changes following electroconvulsive stimulation. Neuroreport 1994; 5:1041–1044
34. Burnham WM, Cottrell GA, Diosy D, Racine RJ: Long-term changes in entorhinal-dentate evoked potentials induced by electroconvulsive shock seizures in rats. Brain Res 1995; 698:180–184
35. Stewart CA, Reid IC: Ketamine prevents ECS-induced synaptic enhancement in rat hippocampus. Neurosci Lett 1994; 178:11–14