Neuron, Vol. 24, 773–776, December, 1999, Copyright ©1999 by Cell Press

The Book of Hebb

Minireview

Terrence J. Sejnowski*
Howard Hughes Medical Institute
The Salk Institute
10010 North Torrey Pines Road
La Jolla, California 92037
Department of Biology
University of California, San Diego
La Jolla, California 92093

It has been 50 years since Donald Hebb published the *Organization of Behavior: A Neuropsychological Theory* in 1949. This book was written at a time when behaviorism was dominant in North American psychology. The approach that Hebb advocated, based on what was then known about the brain, was out of favor among psychologists who believed that only external sensory stimuli and motor responses ought to be included in any explanation of behavior. Most neuroscientists have heard about the "Hebb synapse," but few know why he postulated this learning rule. This is a good time to take a closer look at this book and let Hebb speak for himself.

Hebb was on the faculty of the Psychology Department at McGill University. One of his research interests was the behavioral effects of brain lesions, and he had collaborated with Wilder Penfield, a colleague at McGill, but he was more broadly interested in the development of behavior and learning, which he saw as intimately related. In 1949, much of what we now take for granted about the organization of the nervous system and the properties of neurons was not yet discovered. Hodgkin and Huxley's landmark series of papers on the ionic basis of the action potential would appear in 1952; the classic paper by Fatt and Katz on the quantal theory of synaptic transmission would appear in the same year. Not much was known about the localization of function in the cortex outside primary sensory and motor areas, and Lashley's theory of equipotentiality of the cerebral cortex was still influential.

Most of what was then known about cortical neurons and circuits was based on static pictures of neurons stained with the Golgi technique. Even though the "neuron doctrine" went back to Cajal, conclusive evidence that the neuron was indeed a functional unit awaited the electron microscope in the 1950s and recordings from single cortical neurons in the 1960s. In the introduction to his book, Hebb states that his theory "is evidently a form of connectionism, one of the switchboard variety, though it does not deal in direct connections between the afferent and efferent pathways: not an 'S–R' psychology, if R means a *muscular* response. The connections serve rather to establish autonomous central activities, which then are the basis for further learning" (xix). One of the few figures in the book, reproduced in Figure 1, depicts the connections between area 17 (primary visual cortex) and area 18 (extrastriate visual cortex), and is remarkably modern in including feedback projections and long-range lateral connections within cortical areas as well as feedforward connections. He recognized that single synapses were generally too weak in cortex to cause a postsynaptic neuron to fire a spike and that patterns of converging synaptic inputs were required.

The Motivation for the Hebb Synapse

The central problem that concerned Hebb was the origin of what he believed was relatively autonomous activity in the cerebrum: "... we know practically nothing about what goes on between the arrival of the excitation at a sensory projection area and its later departure from the motor area of the cortex" (xvi). Hebb conjectured that cortical circuits admit self-sustaining activity that reverberated in what he called "cell assemblies." This idea was inspired by evidence for recurrent connections between neighboring cells in the cortex. Although reverberatory activity lasting for up to half a second had been observed by Lorente de Nó, Hebb went further and suggested that such activity in one cortical circuit could through converging projections activate other areas of cortex and lead to a sequence of activations he called a "phase sequence." Although these ideas remain highly speculative, they reflect recent issues such as spike timing and spike synchrony that today are at the forefront of theoretical research on the cortical neural code (Abbott and Sejnowski, 1999).

Hebb needed a way to sustain persistent reverberatory activity (a "trace") in cortical circuits. He proposed that patterns of connections between neurons could sustain reverberatory activity if their strengths could be adjusted by an activity-dependent mechanism for synaptic plasticity that he called a "Neurophysiological Postulate":

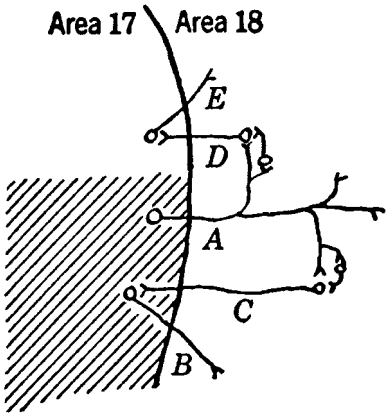

Figure 1. Summary Diagram of Connectivity between Cells in Visual Cortex Taken from Figure 8 in Hebb (1949)

Cells A and B in primary visual cortex (area 17) receive strong excitation from a visual stimulus in their receptive fields (as do other cells in the cross-hatched region). Cells C and D in extrastriate visual cortex (area 18) provide feedback connections to area 17. Cell E does not receive strong visual input in its receptive field but does receive feedback input from horizontal connections within the cortex. Hebb's interpretation of this diagram anticipates the recently discovered modulation of the primary receptive field responses in area 17 from visual stimuli outside the classical receptive field.

*E-mail: terry@salk.edu.

> 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. (62)

This passage is the origin of the "Hebb learning rule" and made Hebb an adjective. Most good ideas have precursors, and earlier versions of Hebb's can be found in books by Jerzy Konorski and even William James, but it was Hebb's version that proved most influential. His words have been interpreted to mean that synaptic plasticity should be based on coincidence detection; that is, strengthening of the synapse should occur when the release of neurotransmitter from a presynaptic terminal coincides with the depolarization of the postsynaptic cell. Evidence for a coincidence detection mechanism has been found in the hippocampus, where long-term potentiation (LTP), discovered there in 1973 by Tim Bliss and Terje Lomo, was shown to be Hebbian (Kelso et al., 1986). LTP of synapses on hippocampal neurons can be elicited by pairing synaptic input with strong depolarizing current, when neither alone produces a long-lasting change, consistent with this interpretation. Furthermore, the induction of LTP at some synapses is controlled by the NMDA receptor, which requires both binding of glutamate and depolarization to allow entry of calcium into the cell. Insofar as the NMDA receptor is a coincidence detector, it might even be called a "Hebb molecule." The only part that Hebb had apparently not gotten quite right was his statement about the firing of cell B, since LTP could still be induced after fast spiking was abolished by blocking active currents in the postsynaptic neuron, suggesting that cooperativity with other synaptic inputs might be needed to depolarize the dendrite sufficiently to open the NMDA receptor. Another issue is that increases in the strength of a synapse from random coincidences will end inexorably in saturation. Hebb suggested that unused synapses might decay, and a form of long-term depression (LTD) induced by low-frequency activity might provide such decay from spontaneous activity in the cortex. However, if synaptic strengths are to encode long-term memories, it is important to have a mechanism for LTD as specific as that for LTP.

Temporally Asymmetric Synaptic Plasticity

Monosynaptic connections between pairs of cells are best examined with dual intracellular recordings in cortical slices. In an experiment designed to test the importance of relative timing of the presynaptic release of neurotransmitter and the postsynaptic activity to LTP, Markram et al. (1997) paired stimulation of cell A either 10 ms before or after spike initiation in cell B. They found reliable LTP when the presynaptic stimulus preceded the postsynaptic spike, but, remarkably, there was LTD when the presynaptic stimulus immediately followed the postsynaptic spike. Similar results have been found for hippocampal neurons grown in culture (Bi and Poo, 1998; Debanne et al., 1998), between retinal axons and neurons in the optic tectum of frogs (Zhang et al., 1998), and in the electrical line organ of weakly electric fish—this is different from the others in that it is of opposite polarity (presynaptic release before the postsynaptic spike causes LTD) (Bell et al., 1997). Thus, this temporally asymmetry in synaptic plasticity is widespread in cerebellar as well as cortical structures. In Figure 2, where the time delay between the synaptic stimulus and the postsynaptic spike was varied over a wide range, the window for plasticity is around $\pm$20 ms and the transition between LTP and LTD occurs within a time difference of a few milliseconds.

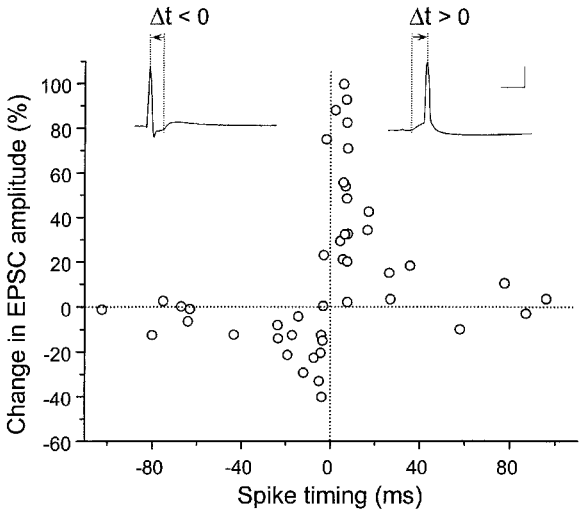

Figure 2. Synaptic Modification in Cultured Hippocampal Neurons

The relative timing of the paired postsynaptic spike and excitatory synaptic inputs (inset) determines whether the subsequent change in the excitatory postsynaptic current (EPSC) increases (when the spike follows the synaptic input) or decreases (when the spike precedes the synaptic input). This is a temporally asymmetric form of Hebbian synaptic plasticity. From Bi and Poo (1998). Calibration: 50 mV, 10 ms.

This temporally asymmetric form of synaptic plasticity has many nice features. First, it solves the problem of balancing LTD and LTP in a particularly elegant way, since chance coincidences should occur about equally with positive and negative relative time delays. Second, when sequences of inputs are repeated in a network of neurons with recurrent excitatory connections, this form of synaptic plasticity will learn the sequence, and the pattern of activity in the network will tend to predict future input. This may occur in the visual cortex where simulations of cortical neurons can become directionally selective when exposed to moving visual stimuli (Rao and Sejnowski, 2000). Similar models have been proposed for neurons in other brain regions, although the temporal window for synaptic plasticity was taken to be 100 ms in the hippocampus (Blum and Abbott, 1996), where there is evidence that the locations of place cells shift to earlier locations in rats running repetitively through a maze (Mehta et al., 1997), and $<$1 ms in a model for learning auditory localization by the relative timing of spikes from two ears (Gerstner et al., 1996).

Is the temporally asymmetric learning algorithm Hebbian? The rapid transition between LTP and LTD at the moment of temporal coincidence does not conform to the traditional view of a Hebbian synapse. Notice that in Hebb's formulation the synapse increases in strength "when an axon of cell A is near enough to excite cell B

and repeatedly or persistently takes part in firing it." For cell A to take part in firing cell B implies causality, not simple coincidence. Thus, the importance of temporal order is implicit in Hebb's formulation. If cell A produces an excitatory event just before cell B fires a spike, then it is likely to have contributed. Hebb did not specify what should happen if cell A fires just after cell B, but weakening is consistent with causality since it is then unlikely for cell A to have caused cell B to fire. The temporally asymmetric learning rule may be more Hebbian than the earlier coincidence version.

For the spike at the soma to influence synapses on distal dendrites, there must be a flow of information from the soma toward the dendrites, which violates the principle of dynamic polarization. This reverse flow of information could not occur without active currents in dendrites, which we now know support exactly the sort of backpropagating action potentials in pyramidal neurons required by the strict form of Hebb's postulate. How the backpropagating spike interacts with the NMDA receptor to produce a knife-edge switch from LTP to LTD is an open research problem.

The temporally asymmetric Hebbian learning rule is equivalent to the temporal difference learning algorithm in reinforcement learning (Rao and Sejnowski, 2000) and can be used to make predictions and implement classical conditioning (Montague and Sejnowski, 1994). The unconditioned stimulus in a classical conditioning experiment must occur before the reward for the stimulus–reward association to occur. This is reflected in the temporal difference learning algorithm by a postsynaptic term that depends on the time derivative of the postsynaptic activity level. The goal is for the synaptic input to predict future reward: if the reward is greater than predicted, the postsynaptic neuron is depolarized and the synapse strengthens, but if the reward is less than predicted, the postsynaptic neuron is hyperpolarized and the synapse decreases in strength. There is evidence in primates that the transient output from dopamine neurons in the ventral tegmental area carries information about the reward predicted from a sensory stimulus (Schultz et al., 1997), and in bees an octopaminergic neuron has a similar role (Montague and Sejnowski, 1994). The temporal window for classical conditioning is several seconds, much longer than the window for LTP/LTD observed at cortical and hippocampal synapses. A circuit of neurons in the basal ganglia and frontal cortex may be needed to extend the computation of temporal differences to such long time intervals. It is surprising to find the same learning algorithm in different types of learning systems in different parts of the brain. This suggests that the temporal order of input stimuli is a useful source of information about causal dependence in many different learning contexts and over a range of time scales.

Spike Timing and Neural Assemblies

Hebb explicitly framed his Neurophysiological Postulate in terms of spikes. Although the traditional coincidence version of the Hebbian learning rule has been applied to many types of neural network models, such as those that use the average firing rates or average membrane potentials of neurons, the temporally asymmetric version of the Hebb rule is most appropriate in models that include spikes. There has been a major effort in the last few years to examine the information carried by single spikes (Rieke et al., 1997) and to analyze models of spiking neurons such as those based on integrate-and-fire processing units or more realistic compartmental models with ion channels based on Hodgkin–Huxley kinetics (Ritz and Sejnowski, 1997). These are being used to explore the computational consequences of the temporally asymmetric Hebbian learning rule.

If relative timing of spikes has a major influence on the strengths of synapses in the cortex, then spike timing is likely to be internally regulated. In particular, local inhibitory interneurons, such as basket cells that can induce rebound spiking in many cortical pyramidal cells, may have an important function in regulating the timing of spikes within a column of neurons (Ritz and Sejnowski, 1997). The relative order of spikes in a population of neurons could also be used to encode information about objects in the world (Hopfield, 1995). In particular, the first neuron in a population to spike in response to a sensory stimulus will have an advantage, since its synapses will be the first to be activated and more likely to be strengthened compared to synapses from other neurons that spike later (Van Rullen et al., 1998; Abbott and Song, 1999). This gets to the core of what assembles cell assemblies and emphasizes how far into the future Hebb's ideas about cell assemblies and phase sequences have endured.

Hebb relied on guesswork and intuition when facts and details were missing. In particular, he focused on synapses as a fundamental computational unit and on activity-dependent synaptic plasticity as a basic operation for both the development of the nervous system and the emergence of higher cognitive functions. Although we now know much more about synapses, and in particular about the conditions under which synaptic strengths can be changed, the links between these changes and cognition remain tenuous (Quartz and Sejnowski, 1997). Hebb recognized that several important levels of organization in the brain above the synapse and neuron were also essential, with neural assemblies having a privileged position. Determining the size of these assemblies, their dynamics, and the extent to which their activities can be linked to perception and cognition remains an exciting research program that will take us well into the next century.

Hebb's book is written in a readable style that has almost disappeared from the scientific literature. It ranges over topics from perception to sleep and from pain to the emotions. He treats each of these issues from the fresh perspective of learning in cell assemblies. In many respects, Hebb could be called the first of the modern cognitive neuroscientists, bringing to bear what was then known about the cognitive, neuroanatomical, neurophysiological, and computational constraints. However, his insights seem to come from a deeper stream and are almost prophetic, making his book part of a much older tradition. I suspect that as we learn more about the brain, we will discover more in Hebb's book that we do not now fully appreciate.

Selected Reading

Abbott, L.F., and Sejnowski, T.J. (1999). Neural Codes and Distributed Representations: Foundations of Neural Computation (Cambridge, MA: MIT Press).

Abbott, L.F., and Song, S. (1999). In Advances in Neural Information Processing Systems, Volume 11, M.S. Kearns, S.A. Solla, and D.A. Cohn, eds. (Cambridge, MA: MIT Press), pp. 69–75.

Bell, C.C., Han, V.Z., Sugawara, Y., and Grant, K. (1997). Nature *387*, 278–281.

Bi, G.-Q., and Poo, M.-M. (1998). J. Neurosci. *18*, 10464–10472.

Blum, K.I., and Abbott, L.F. (1996). Neural Comput. *8*, 85–93.

Debanne, D., Gahwiler, B.H., and Thompson, S.M. (1998). J. Physiol. *507*, 237–247.

Gerstner, W., Kempter, R., van Hemmen, J.L., and Wagner, H. (1996). Nature *383*, 76–78.

Hebb, D.O. (1949). Organization of Behavior: A Neuropsychological Theory (New York: John Wiley and Sons).

Hopfield, J.J. (1995). Nature *376*, 33–36.

Kelso, S.R., Ganong, A.H., and Brown, T.H. (1986). Proc. Natl. Acad. Sci. USA *83*, 5326–5330.

Markram, H., Lubke, J., Frotscher, M., and Sakmann, B. (1997). Science *275*, 213–215.

Mehta, M.R., Barnes, C.A., and McNaughton, B.L. (1997). Proc. Natl. Acad. Sci. USA *94*, 8918–8921.

Montague, P.R., and Sejnowski, T.J. (1994). Learn. Mem. *1*, 1–33.

Quartz, S., and Sejnowski, T.J. (1997). Behav. Brain Sci. *20*, 537–596.

Rao, R.P.N., and Sejnowski, T.J. (2000). In Advances in Neural Information Processing Systems, Volume 12, S.A. Solla, T.K. Leen, and K.-R. Müller, eds. (Cambridge, MA: MIT Press), in press.

Ritz, R., and Sejnowski, T.J. (1997). Curr. Opin. Neurobiol. *7*, 536–546.

Rieke, F.M., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. (1997). Spikes: Exploring the Neural Code (Cambridge, MA: MIT Press).

Schultz, W., Dayan, P., and Montague, P.R. (1997). Science *275*, 1593–1599.

Van Rullen, R., Gautrais, J., Delorme, A., and Thorpe, S. (1998). Biosystems *48*, 229–239.

Zhang, L.I., Tao, H.W., Holt, C.E., Harris, W.A., and Poo, M.-M. (1998). Nature *395*, 37–44.