Invited Review
Canadian Psychology
(2003)

The Once and Future Hebb Synapse

TERRENCE J. SEJNOWSKI
University of California at San Diego

Abstract
The Hebb synapse has become better known than Donald Hebb himself. In this respect he has joined an exclusive club along with the Ising model in condensed matter physics and Parkinson's disease in medicine. This is not to say that Hebb has not made other important contributions, as Peter Milner and Bryan Kolb document, but the Hebb synapse has eclipsed these other achievements. The goal of this essay is to examine how this happened. The Hebb synapse remains a vital organizing concept for both experimental studies and theoretical analysis, as Geoffrey Hinton emphasizes.

I am sometimes asked to identify important advances made by theory and computational modeling in neuroscience. Most would agree that the achievement of Hodgkin and Huxley in modeling the action potential was of seminal importance. Not only did they provide a mechanistic explanation of the action potential that has withstood the test of time, they also outlined a research strategy for explaining even more complex internal properties of neurons that has served us well over the last 50 years (Destexhe & Sejnowski, 2001). It is more difficult to find success stories at the systems level, but Donald Hebb would qualify in my mind. Not only did he make a prediction about the conditions for synaptic plasticity that was subsequently confirmed, he also outlined a framework for building links between neurophysiology and psychology that has become a major research program. I wrote a review of Hebb's 1949 book on the occasion of the 50th anniversary (Sejnowski, 1999) and this essay gives me the opportunity to put recent discoveries in synaptic plasticity into the context of Hebb's research program.

The central problem that Hebb posed in *The Organization of Behavior: A Neuropsychological Theory* was the origin of autonomous activity in the cerebral cortex:

> ... 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. (p. xvi)

Hebb conjectured that cortical circuits admit self-sustaining activity that reverberated in "cell assemblies," inspired by anatomical evidence for recurrent connections between neighbouring cells in the cerebral cortex and reverberatory activity lasting for up to half a second. Hebb further suggested that activity in one cortical circuit could, through converging projections, activate other areas of cortex and lead to a sequence of activations or "phase sequence."

Hebb needed a way to sustain persistent reverberatory activity or "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":

> 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. (p. 62)

These words have been interpreted to mean that the conditions for synaptic plasticity should depend on coincidence detection; that is, strengthening of the synapse should occur when the release of neurotransmitter molecules from a presynaptic terminal coincides with the depolarization of the postsynaptic cell. Instead of being used to develop models of sustained activity in recurrent networks, the first theoretical applications of Hebb's postulate were to models of distributed associative memory in feedforward network models (Hinton & Anderson, 1981). Hebb's postulate admits an alternative, deeper interpretation, which has unfolded over the last five years in a surprising and satisfying way.

HEBB SYNAPSES IN THE BRAIN

Biological evidence for the Hebb rule had to wait for neurobiology to discover conditions that elicited long-term changes in synaptic strength and could be reli-

ably studied at the cellular level. Long-term potentiation (LTP) was discovered in 1973 by Tim Bliss and Terje Lømo in the hippocampus following a high-frequency tetanus. It was rapid in onset, specific for the stimulated synapses and was maintained for many hours. However, it was not until 15 years later that the critical experiment was performed by Tom Brown and his colleagues, who showed that the induction of LTP depended on the postsynaptic neuron and could be prevented by hyperpolarizing it (Kelso, Ganong, & Brown, 1986). Conversely, LTP of synapses on hippocampal neurons can be induced by pairing a weak synaptic input with strong depolarizing current, when neither alone produces a long-lasting change.

The case for Hebb synapses was made even stronger by the discovery that the induction of LTP at some synapses is controlled by the NMDA receptor, whose activation requires the binding of the neurotransmitter glutamate to the receptor and simultaneous depolarization of the postsynaptic neuron. The NMDA receptor is a coincidence detector and could be called a "Hebb molecule." Activation of the NMDA receptor allows calcium ions to pass into the postsynaptic neuron, starting a cascade of biochemical reactions that leads ultimately to strengthening of the synapse.

As Peter Milner points out in his essay, Hebb neglected to include inhibitory influences in his theory and did not suggest specific mechanisms that might lead to decreases in synaptic strength. 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 (Sejnowski, 1977).

During the 1980s, the Hebb synapse was used to implement computational learning algorithms more sophisticated than associative matrix memory, as outlined by Geoffrey Hinton. The beauty of the Hebb synapse is that it only depends on information that is available locally at the synapse, and does not require global information from other neurons in the network such as the error signals required in the error-backpropagation algorithm.

THE NEW HEBB SYNAPSE

Our view of the Hebb synapse has changed following a recent discovery about synaptic plasticity at cortical synapses. The key to making the discovery was the development of new techniques that allow monosynaptic connections between pairs of cells to be examined with dual intracellular recordings. In an experiment designed to test the importance of relative timing of the presynaptic release of neurotransmitter and the postsynaptic activity to LTP, Markram, Lubke, Frotscher, and Sakmann (1997) paired stimulation of a presynaptic neuron 10 ms either before or after initiating a spike in a second postsynaptic neuron. Reliable LTP was observed when the presynaptic stimulus preceded the postsynaptic spike, but, remarkably, there was LTD when the presynaptic stimulus immediately followed the postsynaptic spike. This temporally asymmetry in synaptic plasticity is widespread in the brain. When the time delay between the synaptic stimulus and the postsynaptic spike is varied over a wide range, the window for plasticity is around +/- 20 ms and the transition between LTP and LTD occurs within a time difference of a few milliseconds (Bi & Poo, 1998).

This time-dependent form of synaptic plasticity 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. From a theoretical perspective, this form of synaptic plasticity can learn sequences of spike patterns and will predict future inputs. For example, neurons in a model of visual cortex that incorporated spike-time dependent synaptic plasticity became directionally selective when exposed to moving visual stimuli (Rao & Sejnowski, 2000). Similar models have been proposed for neurons in other brain regions, with temporal windows for synaptic plasticity that were longer – a hundred milliseconds in a model of the hippocampus (Blum & Abbott, 1996), where there is evidence that the locations of place cells shift to earlier locations in rats running repetitively through a maze (Mehta, Barnes, & McNaughton, 1997), or much briefer – less than 1 ms in a model for learning auditory localization by the relative timing of spikes from two ears (Gerstner, Kempter, van Hemmen, & Wagner, 1996)

A closer examination of the Hebb postulate reveals that the spike-time dependent form of synaptic plasticity with a transition from LTP and LTD at zero time delay is closer to Hebb's thinking than the traditional interpretation based on temporal coincidence. Consider his condition for strengthening the synapse: "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, activity in Cell A must have occurred before the spike in Cell B. Thus, the concept of temporal order is implicit in Hebb's formulation. The key principle for triggering synaptic plasticity is causality, not simple coincidence.

Hebb did not specify what should happen if Cell A fires just after Cell B, but weakening is consistent with causality since in this circumstance the spike in Cell A could not have caused Cell B to fire.

Even more surprising was the realization that the new Hebbian learning rule is equivalent to a reinforcement learning algorithm called the temporal difference learning algorithm (Rao & Sejnowski, 2000), which implements classical conditioning (Montague & 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. In monkeys, the transient outputs from dopamine neurons in the ventral tegmental area carry information about the reward predicted from a sensory stimulus (Schultz, Dayan, & Montague, 1997) and in bees, a single octopaminergic neuron has a similar role (Montague, Dayan, Person, & Sejnowski, 1995). It is surprising to find the same learning algorithm popping up in different types of learning systems in different parts of the brain. Perhaps 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.

A similar synaptic mechanism can be used to implement the new version of error-backpropagation that Geoffrey Hinton introduces in his essay. What makes this new theoretical analysis so exciting is that it is based on the microanatomy and physiological mechanisms observed in the cerebral cortex.

NEURAL ASSEMBLIES IN THE BRAIN

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 spike-time dependent version of the Hebb rule requires spikes. There has been a major effort in the last few years to examine the information carried by single spikes (Rieke, Warland, de Ruyter van Steveninck, & Bialek, 1998) 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 (Tiesinga, Fellous, Jose, & Sejnowski, 2002). This leads to the general question of what else spike timing could be used for in the cerebral cortex.

Fries, Reynolds, Rorie, and Desimone (2001) investigated conditions when neurons in visual cortex become synchronized. When a monkey focused attention on a visual stimulus inside the receptive field of a neuron in area V4, there was no change in the firing rate of the neuron, but it became more synchronized with other neurons in V4 firing in the range 30 to 70 Hz. In a computational study where the degree of correlation among the inputs to a model neuron was varied, the output of the neuron was modulated by the correlation strength (Salinas & Sejnowski, 2000). Thus, synchrony in a population of neurons boosts the impact of the spikes by effectively increasing the gain of the downstream postsynaptic neurons. Thus, temporal correlations of spikes can be used to control the flow of information internally within the brain (Salinas & Sejnowski, 2001) and might be used to implement high-level cognitive functions such as attention and expectation. This is different from the view of spike synchrony as a way to represent information in the brain, for example, by binding together the features of an object (Singer & Gray, 1995). Spike rates are still used to represent the sensory input.

Thus, we can begin to see how Hebbian cell assemblies and temporal control of synaptic plasticity might serve as the basis for an integrated theory of attention and memory. Not only can attention regulate the flow of information through an area by synchronizing cells, it can also regulate what is stored in the network at spike-time dependent Hebbian synapses.

CONCLUSION

The context for Hebb's thinking presented in the essays by Peter Milner and Bryan Kolb provide insights into why his approach to behaviour was so fruitful. Rather than formulate his theories in terms of metaphors such as mental chemistry, hydraulics, magnetic fields, telephone switchboards, or information processing, he based his framework on the anatomical structures that were observed in brains. He was far ahead of his time since 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 known. Hodgkin and Huxley's landmark series of papers on the ionic basis of the action potential and the classic paper by Fatt and Katz on the quantal theory of synaptic transmission would appear in 1952. Despite the lack of knowledge of brain mechanisms, Hebb realized that a research program based on bridging the gap from neurophysiology to psychology would be worth pursuing, a field that is now called cognitive neuroscience. The Hebb synapse has matured and continues to be a remarkably vital source of inspiration for both theoretical and empirical studies.

Address correspondence to Terrence J. Sejnowski, Howard Hughes Medical Institute, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037.

Résumé

Aujourd'hui, la synapse de Hebb est mieux connue que Donald Hebb proprement dit. À cet égard il fait partie d'un club très sélect de personnes peu connues mais dont les découvertes ont eu des répercussions incommensurables. Et il suffit de penser au modèle de la matière condensée de Ising en physique et à la maladie de Parkinson en médecine. Ce n'est toutefois pas dire que Hebb n'a pas appporté d'autres contributions importantes, comme Peter Milner et Bryan Kolb le rapportent, mais la synapse de Hebb a éclipsé toutes ses autres réalisations. Le but de cet essai est d'examiner comment cela est arrivé. La synapse de Hebb demeure un concept d'organisation vital tant pour les études expérimentales et les analyses théoriques comme le souligne Geoffrey Hinton.

References

Bi, G.-q., & Poo, M.-m. (1998). Activity-induced synaptic modifications in hippocampal culture: Dependence on spike timing, synaptic strength and cell type. *Journal of Neuroscience, 18,* 10464-10472.

Blum, K. I., & Abbott, L. F. (1996). A model of spatial map formation in the hippocampus of the rat. *Neural Computation, 8,* 85-93.

Destexhe, A., & Sejnowski, T. J., (2001). *Thalamocortical assemblies: How ion channels, single neurons and large-scale networks organize sleep oscillations.* Oxford: Oxford University Press.

Fries, P., Reynolds, J. H., Rorie, A. E., & Desimone, R. (2001). Modulation of oscillatory neuronal synchronization by selective visual attention. *Science, 291,* 1560-1563.

Gerstner, W., Kempter, R., van Hemmen, J. L., & Wagner, H. (1996). A neuronal learning rule for sub-millisecond temporal coding. *Nature, 383,* 76-78.

Hebb, D. O. (1949). *Organization of behavior: A neuropsychological theory.* New York: John Wiley and Sons.

Hinton, G. E., & Anderson, J. A. (Eds.). (1981). *Skeleton filters in the brain, in parallel models of associative memory* (pp. 189-212). Hillsdale, NJ: Lawrence Erlbaum Associates.

Kelso, S. R., Ganong, A. H., & Brown, T. H. (1986). Hebbian synapses in hippocampus. *Proceedings of the National Academy of Science USA, 83,* 5326-5330.

Markram, H., Lubke. J., Frotscher, M., & Sakmann, B. (1997). Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. *Science, 2 75,* 213-215.

Mehta, M. R., Barnes, C. A., & McNaughton, B. L. (1997). Experience-dependent, asymmetric expansion of hippocampal place fields. *Proceedings of the National Academy of Science USA, 94,* 8918-8921.

Montague, P. R., Dayan, P., Person, C., & Sejnowski, T. J. (1995). Bee foraging in uncertain environments using predictive Hebbian learning. *Nature, 3 77,* 725-728.

Montague, P. R., & Sejnowski, T. J. (1994). The predictive brain: Temporal coincidence and temporal order in synaptic learning mechanisms. *Learning and Memory, 1,* 1-33.

Rao, R. P. N., & Sejnowski, T. J. (2000). Predictive sequence learning in recurrent neocortical circuits. In S. A. Solla, T. K. Leen, & K.-R. Müller (Eds.), *Advances in neural information processing systems* (Vol. 12). Cambridge, MA: MIT Press.

Rieke, F. M., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1997). *Spikes: Exploring the neural code.* Cambridge, MA: MIT Press.

Salinas, E., & Sejnowski, T. J. (2000). Impact of correlated synaptic input on output firing rate and variability in simple neuronal models. *Journal of Neuroscience, 20,* 6193-6209.

Salinas, E., & Sejnowski, T. J. (2001). Correlated neuronal activity and the flow of neural information. *Nature Reviews Neuroscience, 2,* 539-550.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275,*1593-9.

Sejnowski, T. J. (1977). Statistical constraints on synaptic plasticity. *Journal of Theoretical Biology, 69,* 385-389.

Sejnowski, T. J. (1999). The book of Hebb. *Neuron, 24*(4), 773-776.

Singer, W., & Gray, C. M. (1995). Visual feature integration and the temporal correlation hypothesis. *Annual Review of Neuroscience. 18,* 555-586.

Tiesinga, P. H. E., Fellous, J.-M., Jose, J. V., & Sejnowski, T. J. (2002). Information transfer in entrained cortical neurons. *Network: Computation in Neural Systems, 13,* 41-66.