Single Channel Theory: A Neuronal Theory of Learning

Richard S. Sutton
(Stanford University)
517 Cowell Lane
Stanford, CA 94305

It has long been believed that learning is somehow due to changes in the connectivity between neurons (synaptic efficacies). But so far adaptive neural network research has found no laws of synaptic efficacy change which can explain the results of learning experiments. Adaptive neural network research using McCulloch-Pitts-type (8) neurons can be divided broadly into two classes (1, 7). One class changes synaptic efficacies based on the occurrence of simultaneous events and is typified by Hebb's theory (increase the efficacy of a synapse when its presynaptic activity coincides with postsynaptic firing (6)). The second class of adaptive neural network research involves theories that change a neuron's afferent (incoming) synaptic efficacies based on the arrival of reinforcement signals *after* the neuron fires. If the reinforcement is positive, the synapse is strengthened, and if the reinforcement is negative, the synapse is weakened. This kind of theory based on subsequent events has been widely used (3, 9, 13). The theory considered in this paper, developed by A. Harry Klopf (7) is also of the subsequent events type. This paper will briefly present Klopf's theory and elaborate on its usefulness as a psychological learning theory. The primary distinguishing characteristic of Klopf's theory as presented here is that it uses only one kind of neuronal signal. Signals that act as reinforcement for neuronal actions are indistinguishable from signals involved in such actions. In deference to this fundamental characteristic, I will refer to the theory as the *single channel theory* of neuronal learning.

Although the single channel theory originated as an adaptive neural network theory, it is also of significance as a theory of learning because of its ability to solve several hard problems in learning theory (according to work in progress (11)). In this short paper one important example will be considered: how the single channel theory can provide mechanism level explanations of classical and operant conditioning, the two basic learning paradigms, as two aspects of a single learning process. Such an ability would be considered an accomplishment for the theory both as a learning theory and as an adaptive neural network theory. Learning theory does not yet understand these two fundamental learning paradigms as much more than procedures for attaining learning (2). Neural network theories have been primarily directed towards explaining one or the other of these learning paradigms and are unable to do both. In this paper my interpretation of Klopf's theory will be presented as a functional (non-physiological) description of neuronal synaptic behavior in the form of three qualitative rules and a mathematical equation. Then two general principles of neuronal behavior in a single channel system will be derived and used to show how the theory is consistent with both classical and operant conditioning.

The Single Channel Theory

In the single channel theory, each neuron uses as its reinforcement the algebraic sum of its inputs from other neurons, with depolarization taken as positive reinforcement and hyperpolarization taken as negative reinforcement. However, the polarization is only effective as reinforcement if it arrives soon after a neuronal firing. This is analogous to the necessity to deliver reinforcement soon after the response to be learned in operant conditioning. Learning experiments indicate that delaying the reinforcement in operant conditioning leads to sharp reductions in learning, with little or no learning if the delay exceeds five seconds (4). Let us call the plot of reinforcement's effectiveness in causing learning versus time delay of reinforcement after response the *reinforcement effectiveness curve*. It will be a consequence of the single channel theory's explanation of classical and operant conditioning that this reinforcement effectiveness curve must be the same curve as the plot of amount of learning versus the conditioned stimulus-unconditioned stimulus interval in classical conditioning. Psychologists consider the effects of these two intervals on learning to be similar (12), but the technical difficulties involved in measuring the detailed characteristics of the reinforcement effectiveness curve in operant conditioning have prevented either confirmation or refutation of this idea. Based on experimental data for the more easily controlled classical conditioning interval (10), the reinforcement effectiveness curve is roughly inverted-U shaped with maximum at 400 msec., and negligible at zero and about 4 seconds (Figure 1). The reinforcement effectiveness curve can be designated as a function of time $E(t)$, where the effectiveness of reinforcement at time t is proportional to $E(t-t0)$, where t0 is the time of the last firing of the neuron.

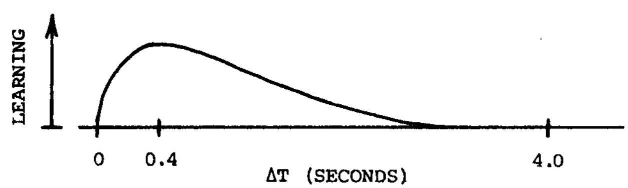

Figure 1: General shape of the reinforcement effectiveness curve

Because of the reinforcement-polarization equivalence there are a couple of potential instability problems (11) in a neural network operating according to the single channel theory. To deal with these Klopf postulated that a synapse which is active (that is providing synaptic input) cannot have its efficacy changed by that input acting as reinforcement. He proposes that in general, whenever a synapse is active, there is a period of time following during which the synapse is ineligible for learning changes. The hypothesized mechanism by which a recently active synapse is kept from undergoing learning changes is called *zerosetting*.

SINGLE CHANNEL THEORY: Sutton

In the qualitative rules that follow a somewhat unrealistic model of the neuron is implicitly used. The most important deviation from reality is that synapses are allowed to change from excitatory to inhibitory and vice versa. In addition, processing by dendritic and other sub-impulse processes are ignored. These properties were chosen for convenience and simplicity in the belief that even if the exact physiological model is not correct, the principles that make the model work should still apply to real nervous systems. The qualitative rules for synaptic change are:

1) Synaptic reinforcement is defined as the albegraic sum of synaptic inputs to the postsynaptic neuron, with depolarization taken as positive and hyperpolarization taken as negative.

2) The synapse is only modified by reinforcement that arrives after a signal is passed through the synapse (i.e., after a presynaptic firing results in a postsynaptic firing). The reinforcement varies in its effectiveness depending on its time of arrival at the postsynaptic neuron according to the reinforcement effectiveness curve and the zerosetting mechanism.

3) The synaptic efficacy is set to the average reinforcement received after a passed signal.

A possible equation for synaptic efficacy change implementing these qualitative rules is:

$$\frac{dG}{dt}(t) \sim [R(t) - G(t)]\ E(t-t0)\ Z(t-t1) \quad \text{(done after firings the synapse contributes to)}$$

where:
- $G(t)$ is synaptic efficacy (determined to a scalar)
- $R(t)$ is instantaneous synaptic reinforcement (algebraic sum of postsynaptic inputs)
- E is the reinforcement effectiveness function
- Z is the zerosetting effectiveness function (a suitable Z function is a block function which is zero for some time and then becomes one thereafter)
- $t0$ is the time of postsynaptic firing
- $t1$ is the time of last presynaptic firing

The Situations that Result in Synaptic Modification

According to qualitative rule three, each of a neuron's afferent synapses will be set proportional to the reinforcement the neuron receives after that synapse passes a signal. To make principles describing the conditions under which a neuron's synapses will undergo these modifications the crucial part here is "after that synapse passes a signal." Modification can occur at a synapse of a neuron only if the neuron fires in response to a synaptic potential generated by that synapse. If the modification is gradual, the synaptic efficacy will be set according to how much reinforcement is received and what proportion of synaptic signal passings are followed by the reinforcement. The amount of modification is independent of reinforcement received when a signal is not passed. Thus, the modification will be just as great if a given reinforcement occurs only when a signal is passed as when the reinforcement occurs every time the presynaptic neuron fires whether or not the postsynaptic neuron fires. Similarly, the synaptic modification will be just as great if the reinforcement comes every time the postsynaptic neuron fires regardless of whether the presynaptic neuron is firing. Thus the situations (reinforcement relationships) in which a synapse will be modified to an equal extent can be divided into three ideal cases:

1. Reinforcement comes only after a certain proportion of the instances of presynaptic firing causing postsynaptic firing.
2. Reinforcement comes only after a certain proportion of the instances of postsynaptic firing, independent of presynaptic firing.
3. Reinforcement comes only after a certain proportion of the instances of presynaptic firing, independent of postsynaptic firing.

In these three cases synaptic efficacy will tend to the same full strength value, but will not, in general, undergo this modification at the same rate measured versus the number of reinforcement instances.

Operant Conditioning and the Contingent Principle

In ideal cases one and two the neuron can be said to be seeking positive reinforcement and avoiding negative reinforcement. In these cases if the reinforcement which follows firing is positive then the synapse will become positive and tend to cause firings. If the reinforcement is negative then the synaptic efficacy will become negative and tend to prevent firings. This behavior can be summarized to some extent in a general descriptive principle:

Contingent Principle: Based on the reinforcement a neuron receives after firings and the synapses which were involved in the firings, the neuron modifies its synapses so that they will cause it to fire when the firing causes an increase in the neuron's expected reinforcement after the firing.

Interpreting stimuli and responses straightforwardly in terms of neuronal firings, this principle leads to an explanation of operant conditioning. The explanation can be illustrated using a simple example of operant conditioning--a hungry rat learning to press a bar to get a pellet of food. The food is known to be a strong positive reinforcer to a hungry rat. In terms of the single channel theory this means that the food causes much more excitation in the rat's brain than inhibition. The sight of the bar is the conditioned stimulus (CS) and the movements in bar pressing are the conditioned response (CR). If the rat performs the CR (bar pressing movements) in response to the CS (sight of the bar), then it gets reinforcement (the food), and subsequently tends to perform the CR to the CS more often and more efficiently. According to the single channel theory those neurons responsible for the CR (bar pressing movements) "learn," as summarized in the contingent principle, that if they fire in response to the CS (presumably some signal indicating the CS is accessible to these neurons at some of their synapses) then they will receive positive reinforcement (the excitation distributed to the brain when the food is received). The neurons responsible for the CR will "learn" by making their synapses whose presynaptic

SINGLE CHANNEL THEORY: Sutton

activity signals the CS more positively effective in causing the neuron to fire. Thus, the CR will be more likely to occur in response to the CS.

Classical Conditioning and the Predictive Principle

The behavior in ideal cases one and two has been summarized in the contingent principle and used to explain operant conditioning. In ideal case three there is a very different situation from that in the other two cases. Here the reinforcement received by the neuron is independent of whether it fires or not. This case causes us to use a second general principle in explaining and understanding the behavior of neurons operating according to the single channel theory:

Predictive Principle: If a synapse's activity predicts (frequently precedes) the arrival of reinforcement at the neuron, then that activity will come to have an effect on the neuron similar to that of the reinforcement.

If activity in some synapses predicts the arrival of positive reinforcement, then the synapses will become positive, and if the predicted reinforcement is negative, then the synapses will become negative.

The general classical conditioning procedure consists of presenting a neutral CS, one that does not cause a particular response other than orienting responses, followed by an unconditioned stimulus (UCS) which reflexively causes an unconditioned response (UCR). After a number of such pairings of the CS and the UCS-UCR, the CS assumes the power to evoke a response of its own which closely resembles the UCR or some part of it. Classical conditioning is easily explained using the predictive principle. Consider the neurons responsible for the UCR. By definition, these are caused to fire by the UCS. Thus the UCS must cause them to be excited, and thus the UCS is positive reinforcement to these neurons. If these neurons have access to a signal at some of their synapses that indicates the CS, then these synapses' presynaptic activity will predict (frequently precede, by experimental design) the arrival of the positively reinforcing UCS. Thus, by the predictive principle, these synapses will become positive and tend to cause the neurons responsible for the UCR to fire when the CS occurs. Referring back to ideal learning case three, whence the predictive principle was derived, it is apparent that for the neurons responsible for the UCR to undergo learning changes they must sometime fire in response to the CS. This is the only condition for the synapses signalling the CS to undergo learning changes that is not explicitly fulfilled in the classical conditioning paradigm. This condition will also be satisfied if by chance some of the synapses signalling the CS already happen to be slightly positive or if one of the neuron's occasional background firings occurs while the CS is on and these synapses are thus presynaptically active. It is very likely that the synaptic learning changes will occur in these situations because the nonlearned state is unstable. Once the synapses become slightly positive they will tend to cause the neuron to fire more, thus increasing the opportunity for synaptic modification. The occasional chance firing of the neuron in response to an arbitrary input is probably not unlikely. Neurons are constantly firing at an average rate on the order of one to ten times per second (in cat visual cortex (5)), so that if the CS's and its signalling synapses' activity lasts very long such a coincidence is very likely in some of the neurons responsible for the UCR. In order for a synapse to undergo learning changes, its presynaptic activity must precede reinforcement and the activity must result in the postsynaptic neuron's firing. Classical conditioning comes about in what is in some sense a degenerate case. In classical conditioning the CS's synapses' activity is a good predictor of the coming reinforcement (the UCR) whether or not the neuron fires to it. Thus it is also a good predictor when the neuron does fire to it, and learning occurs.

Although neuron learning behavior has been divided into three ideal cases, and two learning principles have been derived, the single channel theory is not a multi-factor theory of learning. The qualitative rules make it clear that there is only one learning process. The cases and principles are all just different aspects of this process.

In addition to providing a single process explanation of classical and operant conditioning, the single channel theory also appears to be able to solve several other hard problems in learning theory. The theory has been extended to explain classical and operant conditioning interaction in secondary reinforcement and unmotivated learning in sensory preconditioning and latent learning (11). In addition, the single channel theory has been mathematically formalized and its basic learning capabilities have been tentatively verified in computer simulation of small networks of neurons (11). I am presently attempting to extend the theory to explain certain other aspects of learning theory including avoidance learning, conditioned emotional response, overshadowing, and other expectation phenomena.

References

(1) Arbib, M. A., (1975), From automata theory to brain theory, *Int. J. Man-Machine Studies*, **7**: 286.

(2) Bolles, R. C., (1975), *Learning Theory*, Holt, Rinehart and Winston, 159-166.

(3) Farley, B. G., and Clark, W. A., (1954), Simulation of self organizing system by a digital computer, *IRE Trans on Information Theory*, **PGIT-4**: 76-84.

(4) Grice, G. R., (1948), The relation of secondary reinforcement to delayed reward in visual discrimination learning, *J. Exp. Psychol.*, **38**: 1-16.

SINGLE CHANNEL THEORY: Sutton

(5) Griffith, J. S., (1971), *Mathematical Neurobiology*, Academic Press, London, 22.

(6) Hebb, D. O., (1949), *The Organization of Behavior*, John Wiley and Sons, New York.

(7) Klopf, A. H., (1972), *Brain Function and Adaptive Systems - A Heterostatic Theory*, Air Force Cambridge Research Laboratories Report.

(8) McCulloch, W. S., and Pitts, W. H., (1943), A logical calculus of the ideas immanent in nervous activity, *Bull. Math. Biophys.*, 5: 115.

(9) Rosenblatt, F., (1957), *The Perceptron: A Perceiving and Recognizing Automaton, Project PARA*, Cornell Aeronautical Laboratory Report 85-460-1.

(10) Russell, I. S., (1966), Animal learning and memory, in Derek Richter, ed., *Aspects of Learning and Memory*, Basic Books, New York, 136.

(11) Sutton, R. S., (1977), Learning theory support for a single channel theory of the brain, unpublished draft.

(12) Tarpy, R. M., (1975), *Basic Principles of Learning*, Scott, Foresman and Company, Glenview, Illinois, 54.

(13) Widrow, B., (1962), Generalization and information storage in networks of adaline neurons, in Marshall C. Yovits, George T. Jacobi and Gordon D. Goldstein, eds., *Self Organizing Systems*, Spartan Books, Washington, DC, 435-461.