6/27/78

LEARNING THEORY SUPPORT FOR A SINGLE CHANNEL THEORY OF THE BRAIN

Richard S. Sutton

LEARNING THEORY SUPPORT FOR A SINGLE CHANNEL THEORY OF THE BRAIN

TABLE OF CONTENTS

SECTION PAGE

LEARNING THEORY SUPPORT FOR A SINGLE CHANNEL THEORY OF THE BRAIN

1.0 INTRODUCTION

A common way to develop general theories of the brain is to theorize about the neuron as the fundamental building block. Frequently the neuron is modeled as an input-summing threshold device and learning is proposed to reside in the connections with other such elements. The question has always been how to change the efficacy of the connections as a function of past experience so that the network of neurons has brain-like learning properties. This paper will support a theory of neuronal operation developed by A. Harry Klopf by citing the evidence of psychology's study of learning. For instance, this paper will show how the theory provides a unique and highly satisfactory unification of classical and operant conditioning, the two major learning paradigms, as two aspects of the same basic learning process.

Klopf's theory is one in which the synaptic efficacy undergoes changes based on the arrival of reinforcement after neuronal firing. This is in contrast to theories such as D. O. Hebb's (Hebb, 1949) that change the efficacy of synapses upon the occurrence of simultaneous events (in Hebb's case the simultaneous events are presynaptic and postsynaptic firing). The primary distinguishing characteristic of Klopf's theory as presented it here is that it uses only one kind of neuronal signal. Signals that act as reinforcement for a neuronal action are indistinguishable from the signals involved in such actions. In deference to this fundamental and unique characteristic, I will refer to the theory as the single channel theory of neuronal learning.

An implicit model of much of stimulus-response theory has been that of a central reinforcement mechanism which can change the connections used just prior to its activation. The neural modelers took up on this idea and developed the perceptron and various perceptron based systems. In the perceptron as in the central reinforcement mechanism model a single evaluation of performance is used throughout the system to change connections. Because the reinforcement mechanism affects the entire system these theories are said to use global reinforcement. Some theorists now believe that some principle of local reinforcement is necessary to explain all of the brain's experimentally demonstratable capabilities. Local reinforcement refers to the use of different evaluations of the results of an action in different parts of the system. A major advantage of a local reinforcement theory is that it can potentially learn many things at once. Each of the subsystems with different evaluation measures can learn seperately of and in parallel with the others. In theory, more can be learned this way in a given amount of time. The system might evaluate an action as being executed successfully but as having an undesirable result. In this case, one should reward the execution and punish the the decision to perform the action in that particular situation. Such divisions in the scope of reinforcement distribution are by definition impossible in a global reinforcement system.

The single channel theory of neuronal learning proposes that the brain is a local reinforcement system. In this theory, each neuronal element has as its reinforcement the algabraic sum of its inputs from other neurons, with depolarization taken as positive reinforcement and hyperpolarization taken as negative reinforcement. In principle then, each neuron can have its own measure of success, and be striving after its own goals. However, the reinforcement measures can be effectively the same in different neurons to the extent that they are excited or inhibited from the same sources. In particular, it is hypothesized by this theory that the neurons of the brain are controlled to attain otherwise arbitrary objectives like food, shelter, and physical integrity by occasionally distributing more or less global excitation and inhibition.

As a local reinforcement theory, the single channel theory has the potential for learning things that are impossible for a global reinforcement system. A global reinforcement system can not notice and learn from globally neutral events. Thus, it can not learn an association between paired neutral events nor can it learn how to accomplish a goal if the reinforcement system for that goal is at present sated and unmotivated with respect to that goal. Yet all these things the experimental study of learning has shown that animals can and do do very frequently and naturally. Neurons operating according to the single channel theory clearly can notice relationships between globally neutral stimuli because since all stimuli excite some neurons, all stimuli reinforce some neurons, and these reinforced neurons will learn about what brings them reinforcement. However, it remains to be seen whether the whole system will be able to make constructive use of the relationship noticed by the few neurons reinforced by the stimuli in order to accomplish its system goals.

For the single channel theory of neuronal learning a very serious question of constructability must be faced. It seems a little preposterous to propose that usual neural excitation and inhibition can be used as reinforcement to the neuron and then to go on theorizing about all the marvelous advantages of this setup without carefully checking to see if the idea is even mathematically sound and constructable. Thus, to insure mathematical soundness and to check the major, simpler predictions about the behavior of networks of neurons working according to the single channel theory, computer simulations have been developed. This work, including detailed mathematical formulation of the neuronal equations for synaptic change is also reported on in this paper.

2.0 THE SINGLE CHANNEL THEORY

2.1 Making A Single Channel Theory Work

The initial objection to an input-reinforcement equivalence is that it seems the learning system would not tend to learn anything useful. It seems it would be easiest for the neurons to pass signals amongst themselves and relatively ignore trying to manipulate the environment. Circular loops of neurons such as those shown in schematic (figure 1) are known to be not uncommon in the brain. If neuron A (refer to figure 2) increased synapse C when firing in response to C's passing a signal from B and positive reinforcement following, and if the same held for neuron B and synapse D (initially by chance), then such structures would quickly turn into positive feedback systems and the synapses would become large. This kind of learning does not seem very useful. The solution proposed by Klopf is to specify that a synapse which is active (that is passing a signal) can not have its efficacy changed by that signal acting as reinforcement. He proposes that in general, when a synapse is active, there is a period of time following during which the synapse is ineligible for learning changes. Thus, for effective learning to occur, a neuron must have two afferent synapses. One to carry the signal that the neuron fires in response to, and one to carry the reinforcement to evaluate the firing to the signal and change the synapse (figure 3). In order for the learning to occur, the reinforcement must come some time after the input signal occurred, and specifically, the input signal can not be occurring

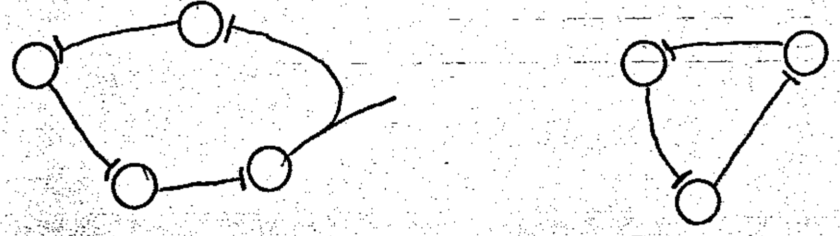


Figure 1.

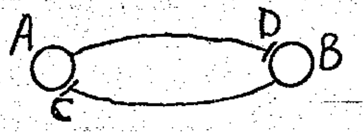


Figure 2.

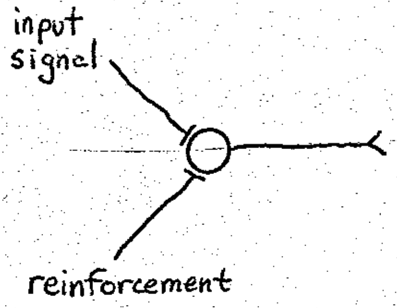


Figure 3.

when the reinforcement arrives. The hypothesized mechanism by which a recently active synapse is kept from undergoing learning changes is called zerosetting.

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 reinforcment in operant conditioning leads to sharp reductions in learning, with little or no learning if the delay exceeds five seconds (Grice, 1948). 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 (Tarpy, 1975), but the technical dificulties 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 (Russell, 1966), the reinforcement effectiveness curve is roughly inverted-U shaped with maximum at 400 msec., and negligible at zero and about 4 seconds (figure 4). The reinforcement effectiveness curve can be designated as a function of time E(t), where the effactiveness of reinforcement at time t is proportional to E(t-tO), where tO is the time of the last firing of the neuron.

The general intent of zerosetting and the reinforcement effectiveness function is to force the neurons to have to learn about manipulating the environment. Consider the diagram of the feedback loops available to a neuron in figure 5. The presumption is that for a neuron to learn along a reinforcement loop of sufficient delay to escape zerosetting, the loop will have to involve the learning of something potentially useful and significant for the system. Typically, this may mean something about the external or internal environment, but it may also mean a specially set up neural loop, such as the circuit of Papez in the limbic system may be.

2.2 Qualitative Synaptic Change Rules

Real neurons are believed to have their synapses fixed whether they are excitatory or inhibitory and thus learning changes probably only involve the extent to which the synapses are excitatory or inhibitory. For simplicity of the discussion and the formal model, the synapses will be allowed to be both

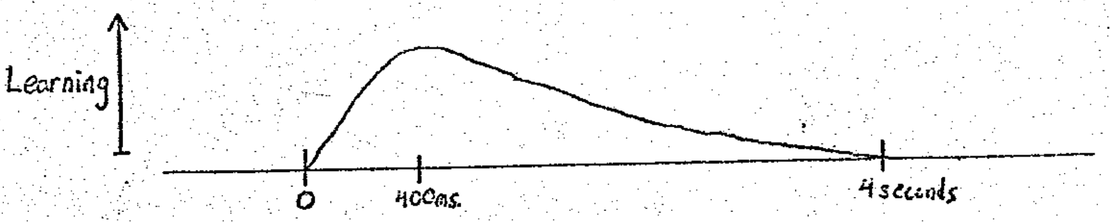


Figure 4. Time interval between response and reinforcement deliverance.

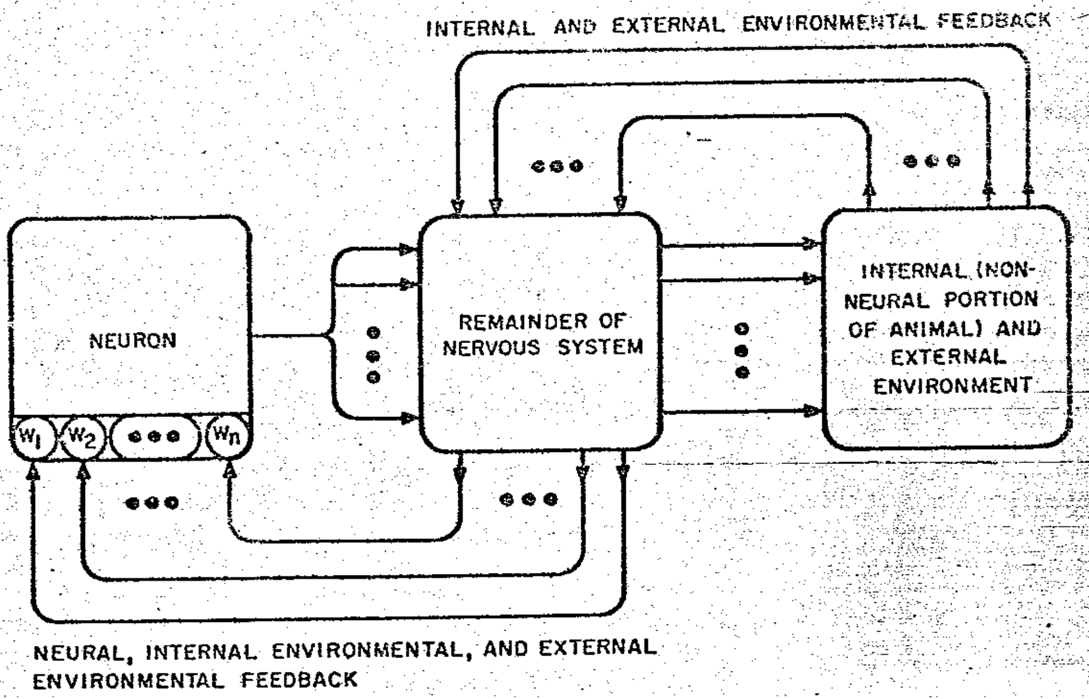

INTERNAL AND EXTERNAL ENVIRONMENTAL FEEDBACK
NEURON
w_1
w_2
w_n
REMAINDER OF NERVOUS SYSTEM
INTERNAL (NON-NEURAL PORTION OF ANIMAL) AND EXTERNAL ENVIRONMENT
NEURAL, INTERNAL ENVIRONMENTAL, AND EXTERNAL ENVIRONMENTAL FEEDBACK

Figure 5 Relationship of a Neuron to the Remainder of the Nervous System and to the Internal and External Environments. w_1, w_2, ..., w_n = variable synaptic transmittances

(From Klopf, 1972)

excitatory and inhibitory at different times. The synaptic efficacy can then be thought of as a single number which can, through learning, change its value and even its sign. A positive synapse will mean it is excitatory, causing depolarization, and a negative synapse is inhibitory, causing hyperpolarization. This sign changing capability should not severely affect the generality of our conclusions. These properties were chosen primarily 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.

In addition, for the following set of synaptic change rules, we will speak as if each firing occurred in isolation rather than being mixed in as one of a series of firings, and we will speak as if a synapse is either involved in (causes) a firing or is not rather than this actually being a continuum. Later in this paper when these qualitative rules are formalized mathematically they will be extended and refined to cover these cases in complete generality. The qualitative learning rules for synaptic change can now be stated:

1) Synaptic reinforcement is defined as the algebraic 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.

2.3 What The Neuronal Elements Learn

The synapse is the natural unit for describing how changes due to learning occur. But the natural unit for describing what is learned is at a higher level. What is learned at the level of the individual neuronal element will be described next. The properties that we establish here about what the neuron learns will then be used to show what a neuronal network learns.

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 or not. 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. 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.

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 synapse will become negative.

The two ways of describing what the neurons learn (the general principles just stated) correspond exactly to the two major conditioning paradigms of learning theory - operant and classical conditioning. According to the single channel learning theory, neurons learning to fire so as to get more reinforcement is the kind of learning involved in operant conditioning and neurons learning to fire to predictors of reinforcement is the kind of learning responsible for classical conditioning. Not only are operant and classical conditioning unified this way as two aspects of the same common learning process, but secondary reinforcement and sensory preconditioning also are explained with no further embellishment of the learning rules. A description of each of these phenomena of learning theory and their explanation by the single channel theory follows.

3.0 AN EXPLANATION OF SOME FORMS OF LEARNING

3.1 The Role Of The Predictive Principle In The Nervous System

By the predictive principle we propose that the neurons of the brain are learning to have predictors of stimuli have the same effect on them as the stimuli themselves. From this principle and the assumption that responses and stimuli are caused by neuronal firings and cause neuronal firings respectively, we see that if a stimulus causes a response, then a second stimulus which predicts the first will also come to tend to cause the response. This is exactly what is experimentally observed as classical conditioning. 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 a unconditioned stimulus (UCS) which reflexively causes a 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 signaling 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 (Griffith, 1971)), so that if the CS's and its signaling 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 synapse's 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. Some of classical conditioning's more detailed characteristics such as when it works and the character of the conditioned response turn out to fit in very well with this explanation (Klopf, 1972).

Learning by the predictive principle is not dependent on the actions of the neurons or of the organism for the occurrence of subsequent reinforcement. Thus, it seems that this learning should occur even when the predicted stimulus does not have any external response associated with it. If the predicted stimulus delivers excitation to some neurons, then a predicting stimulus should also come to excit those neurons (if they have access to signals of the predicting stimulus). Interestingly enough, just such learning has been found experimentally to occur (Coppock, 1958) and has caused difficulty for many of the learning theorists (Tarpy, 1975). Experiments show that animals will automatically generalize from responding to particular stimuli to responding to associated stimuli, even if the association occurs before the response is learned. The forming of the association before response learning is called sensory preconditioning. There are two learning stages and two testing stages in the sensory preconditioning experiment. Here is an example:

Learning 1: Repeatedly present a tone and then a light, with time interval 400 ms.

Learning 2: Repeatedly present the light, then, 400 ms. later, a foot shock sufficient to get a leg flexion response.

Testing 1: Present light. Since learning 2 should have classically conditioned leg flexion to the light, the animal should show a CR resembling leg flexion.

Testing 2: Present tone. The animal also shows the leg flexion CR, just as if leg flexion had been classically conditioned to the tone as well as to the light.

Since there was no apparent reinforcement associated with the tone, a central reinforcement mechanism theory is hard pressed to explain why any reponse could have been learned to it.

The model neurons explain sensory preconditioning because they learn, in the same way that they learn by classical conditioning, to have the signal that is a predictor of a neuron affecting stimulus, come to have the same effects as that stimulus. Sensory preconditioning points out that even when this does not have an effect on the animals behavior, as when he does not yet have any particular response to the predicted stimulus, the prediction association is still learned and the synaptic connections made.

3.1.1 Global Reinforcement -

According to the theory, one source of reinforcement must be the sensory input channels, as these undoubtedly carry excitation to parts of the brain. But it is clear that man and animals are motivated by other things than arbitrary stimulation. Somehow the brain has been structured, apparently by its genetic inheritance, so that it learns to maintain itself and its species by eating and drinking, avoiding tissue damage, sex and childrearing, etc. Thus, the theory must propose that certain sensory events are specially wired up to cause to cause either general excitation or general inhibition over relatively large areas of the brain. This is effectively a kind of global reinforcement which is, as we observed earlier, a necessary part of any useful or directable local reinforcement system.

3.1.2 Secondary Reinforcement -

It has long been realized that the brain's reinforcement mechanism is not simple. The model of a central mechanism to deliver reinforcement to the entire brain and which sends reward signals when we eat and punishment signals when we get hurt must soon be modified to take account of the prominence of secondary reinforcement. Secondary reinforcement refers to the fact that stimuli associated with reinforcement take on reinforcing properties themselves. If the deliverence of food and a clicking sound are repeatedly paired, an animal such as the rat will learn to do work such as bar pressing just to hear the click, even if it is no longer paired with food (in which case the effect would extinguish before very long). Any plausible reinforcement mechanism must be flexible enough to allow an arbitrary stimulus to take on reinforcing properties when paired with reinforcement. The immediate way to modify the central reinforcement mechanism theory is to claim that the central mechanism learns by classical conditioning that certain stimuli are associated with reinforcement and sends out reinforcement in response to them also. In a local reinforcement theory the role of central mechanisms is not so prominent and in the present theory secondary reinforcement is handled very naturally without sending the secondary reinforcing stimulus signals to and from a central mechanism.

The single channel theory's explanation of secondary reinforcement is immediate from the predictive principle and the synaptic input-reinforcement equivalence. All neurons which receive a reinforcement (which is just another nonzero input) and which have access to a predictor of that reinforcement, will learn to have the predicting input have the same effect as the predicted input - it will become a reinforcer. If the learned secondary reinforcement is positive this learning will cause some of the neurons to fire in response to the input signal, thus increasing the number of neurons that have access to the signal and further spreading the range of the secondary reinforcement effect. In this case, the effective scope of the secondary reinforcement will be precisely that of the original reinforcement that it predicts. Thus, a signal that predicts the arrival of food tends to reinforce as much of the brains neurons that were reinforced by food, which is presumed to be a significant portion to provide control of responses in the hungry animal. The secondary reinforcement is in every way equivalent to the original reinforcement. Thus, if another signal is a predictor of the predictor of reinforcement, then it too will take on reinforcing properties. In principle this chaining can go back for any length and thus provides an explanation why, under certain conditions, animals can learn tasks that involve a delay between response and reinforcement, or between CS and UCS in classical conditioning, of more than the few seconds that neurons are allowed to look at reinforcement after firing. The conditions under which this kind of learning with delayed reinforcement are possible are those in which there are secondary reinforcement cues available to the animal. This means that there is some signal after response performance that occurs within a few seconds and optimally 400 ms. after, which indicates that the reinforcement is coming. In natural situations or in normal learning experiments the use of this kind of reinforcement is the rule rather than the exception. A rat is rewarded for pressing the bar not so much by the eventual presence of food in its stomach, but more by the sight of the food being in the hamper or by the sounds that signify that it has been put there.

3.2 The Role Of The Contingent Principle In The Nervous System

By the contingent principle we propose that the neurons of the brain learn to fire in general and to fire to certain inputs in order to maximize reinforcement that is dependent on their firing. Interpreting stimuli and responses straightforwardly in terms of neuronal firings, this 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 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.

This learning of the neurons by the contingent principle of when to fire and whether to fire in order to get their reinforcement is not dependent on the reinforcement being global reinforcement. Even a "neutral" sensory stimulus excites some neurons of the brain and thus these will learn when to fire in order to get this positive reinforcement. If the stimulus is truely neutral then an equal number of neurons will be inhibited by the stimulus and will try to fire so as to prevent its occurrence. Thus, parts of the brain will learn about the response contingencies by the contingent principle just as we have shown that parts of the brain learn about prediction relationships among neutral stimuli by the predictive principle.

There is strong experimental evidence that nervous systems do learn about the response contingencies of neutral stimuli. The general phenomena is called latent learning. A good early example is the famous case of Blodgett's rats. Blodgett taught two groups of rats to go through a maze to get food. One of them, the experimental group, had first been familiarized with the maze by allowing them to completely explore it, without their being any food contained in it. He found that the rats that had already explored the maze learned much faster than the control group who had just as many reinforced trials. Apparently, even though there was no reward, the rats had learned something about the maze by exploring it beforehand.

According to the single channel theory of the brain, neurons in parts of the rats' brains learned what responses would get what stimuli, what turns in the maze would get to what place in the maze, as explained earlier by the contingent principle. Experimentally, we see that somehow the rats were able to use this previously learned information once it became globally useful. In order to explain this ability to use the learning about neutral stimuli in latent learning, a third general principle of neuron behavior will have to be proposed:

Activity Change Principle: Neurons have changes in their frequency of firing based on the reinforcement that arrives after firing. If the reinforcement is positive, the neuron tends to fire more often, if it's negative, the neuron fires less often.

Now we can explain latent learning. Consider that there have been formed two groups of neurons, one gets positively reinforced by a stimulus and which has learned when to fire to get it, and one which is negatively reinforced by the stimulus and which has learned how to avoid it. The two have counteracting effects resulting in no overt response tendency. Now assume the stimulus occurs and is followed by positive reinforcement (this corresponds to when the latent learning becomes useful - now there is food in the maze). Only the neurons positively reinforced (excited) by the stimulus will fire when the stimulus occurs. Thus, when the positive reinforcement follows, these will get an increased tendency to fire subsequently while the neurons that have learned to prevent the stimulus will not. Thus, the neurons trying to get the stimulus will dominate by their increased firing and tend to control the organism's actions.

To the extent that the activity change principle is redundant with the contingent principle it has been included in the mathematical formalization of the neuron. However, I have not particularly tried to implement the activity change principle yet in the computer simulations.

4.0 MATHEMATICAL FORMULATION

4.1 The Basic Form Of The Equations

Rule number 3 of our qualitative rules for synaptic efficacy change pointed out that in repeated identical trials of presynaptic firing causing postsynaptic firing and reinforcement following, the synapse tends to go to a full strength that is proportional to the average reinforcement after firing. Thus the synaptic efficacy G (for gamma), should be set proportional to average reinforcement after a firing, but only counting those firings to which the synapse contributed to. This could be done by defining reinforcement after a firing as the sum of all the inputs after the firing, but the problem with this formulation is that it does not allow the synaptic efficacies to change smoothly with time. Our equations can not use equations with terms such as all reinforcement after a firing which imply waiting several seconds over which to add up instantaneous reinforcement values to find total excess. Instead, it behooves us to develop equations using only instantaneous measures of reinforcement. This is not a serious problem because we can find a relation between instantaneous reinforcement and what synaptic efficacy should be - G should be proportional to average instantaneous reinforcement. This relation must hold to insure that G is proportional to excess reinforcement which is proportional to average instantaneous reinforcement.

What is needed is an incremental or derivitive equation for G that can be shown to have the desired tendencies. In particular, we would like the increment or derivitive to be zero when the ideal value for efficacy is reached. Thus, I propose we consider:

1) $$\frac{dG}{dt}(t) \sim R(t)-KG$$ (done after firings the synapse contributes to)

where: R(t) is instantaneous reinforcement
G is synaptic efficacy
K is an arbitrary constant

This is zero when G bears a certain proportionality to instantaneous reinforcement and the sum effect of the equation is zero (they cancel) if G bears that proportionality to average instantaneous reinforcement (assuming instantaneous reinforcement is averaged and the above equation is applied in the same way). This will be the basic form of our synaptic change equations. Note that while G's derivitive is given as a function of time while G itself is not. This is done to emphasize that G normally changes slowly in relation to the time course of the reinforcement effectiveness curve. Thus, G can be viewed as a constant for the duration of a firing and the subsequent effective reinforcement. The <u>change</u> in G, on the other hand, changes very quickly dependent on the momentary reinforcement being received by the neuron.

Zerosetting can be added to this equation by constructing a function for the zerosetting mechanisms effectiveness that is zero when the mechanism is fully effective and one when totally ineffective. The derivitive equation can then be multiplied by this function so that it acts as a weighting of the significance of the present situation:

$$2)\quad \frac{dG}{dt}(t) \sim [R(t)-KG]\, Z(t-t1) \qquad \text{(done after firings the synapse contributes to)}$$

where: Z is the zerosetting effectiveness function
t1 is the time of last presynaptic firing

A suitable Z function is a block function which is zero for some time and then becomes one thereafter.

When zerosetting was discussed earlier in this paper it was determined that reinforcement should be effective after a response according to an inverted-U shaped reinforcement effectiveness curve. A firing is the neural equivalent of a response so the effectiveness of reinforcement as it comes in after a firing should be weighted by this curve:

$$3)\quad \frac{dG}{dt}(t) \sim [R(t)-KG]\, Z(t-t1)\, L(t-t0) \qquad \text{(done after firings the synapse contributes to)}$$

where: L is the reinforcement effectiveness function
t0 is the time of postsynaptic firing

This reinforcement effectiveness function E(t) is only known experimentally and thus has no mathematical formulation yet. This function forms the basis for a reinforcement effectiveness function which will be developed in the next section (zerosetting, in a sense, also weights reinforcement effectiveness but is not considered part of the reinforcement function because it introduces unnecessary intractable mathematical dificulties).

4.2 Refining The Formulation And Extending To Many Firings

4.2.1 The Weighting Effectiveness Of Reinforcement Function -

Thus far we have considered weighting reinforcement's efectiveness by an inverted-U shaped curve traced out by a function L(t) of the time since postsynaptic firing. In this section the reinforcement effectiveness function will be refined in several ways. First, instead of adding the proviso to only apply the synaptic change equation when the synapse contributes to the firing, this will be included in the reinforcement effectiveness function. The binary concept of contributing or not contributing will be generalized to a continuous measure of the synapse's contribution. After this, the reinforcement effectiveness function will be generalized to the case of many firings in the neuron and the reinforcement effectiveness function L(t), will be given a mathematical embodiment.

When a neuron fires a single time each of its synapses may have contributed to that firing to different degrees and thus should be changed to different degrees. Up to now we have been assuming that either a synapse contributes or does not, with no inbetween. This appears to be an artificial division. Our crucial measure of contribution is the time between presynaptic firing and postsynaptic firing. Some arbitrary time criteria could be chosen to divide the synapses into contributers and noncontributers, but it is more natural to use a more continuous measure of degree of contribution. A simple model is that of exponential decay:

$$4)\quad C = e^{-(t0-t2)/T}$$

where: C is the contribution of this synapse
t0 is time of the postsynaptic firing
t2 is time of the last presynaptic firing before t0
T is the membrane potential decay time constant

Since the time constant is that of the membrane potential's decay, this C will be proportional to how much of the last impulse passed by the synapse was still around when the postsynaptic neuron fired. Now the effectiveness of reinforcement funtion can be refined to take account of the size of the synapse's contribution to the firing:

$$5)\quad E(t) \sim C\ L(t-t0)$$

where L(t) is the reinforcement effectiveness function to be used this way in the synaptic change equation:

$$6)\quad \frac{dG}{dt} \sim [R(t)-KG]\ L(t)\ Z(t-t1)$$

which now no longer needs the after firings the synapse contributed to proviso.

At this point soma terminology should be revised and clarified. Let us reserve the symbol L(t) for the function producing the single-humped inverted-U shaped curve previously called the reinforcement effectiveness curve. And let us use the symbol E(t) and the name reinforcement effectiveness function for a synapse specific history of the effectiveness of postsynaptic reinforcement in changing that synapse. Thus, the reinforcement effectiveness function E(t) may consist of several humps of the shape traced by L(t), whose onset corresponds to firings in the postsynaptic neuron. The diagram in figure 6 of a synapse's reinforcement effectiveness function E(t), shows the result of two firings (they are assumed to be seperated by enough time to be considered two seperate firings). The firings cause temporary increases in the E function that are of the same shape - that of the function L(t), but which are of different heights dependent on how closely postsynaptic firing followed presynaptic firing.

Now we can consider what should happen to the E function if there are two firings occurring close enough together that the E function has not yet become zero from the first one when the second occurs. Consider the E functions as they would be if each firing had occurred alone. Then we could apply our equation once for each firing:

$$7)\quad \frac{dG}{dt}(t) \sim [R(t)-KG]\ Z(t-t1)\ E1(t) + [R(t)-KG]\ Z(t-t1)\ E2(t)$$

where E1 is the first firing's seperate E function
and E2 is the second firing's seperate E function.

But since this is equivalent to:

$$8)\quad \frac{dG}{dt}(t) \sim [R(t)-KG]\ Z(t-t1)\ [E1(t)+E2(t)]$$

the equation can just be used once once with an E function that is the sum of each firing's E function seperately. This concept is graphed in figure 7. The concept is easily generalized to any

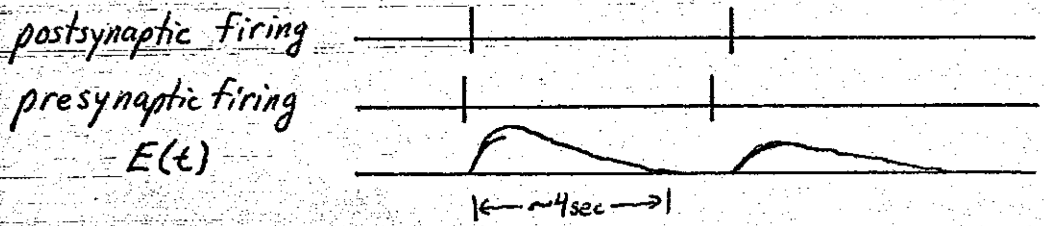


Figure 6. Two postsynaptic firings with different times since the last presynaptic firing. The E function humps are of the same shape but different heights.

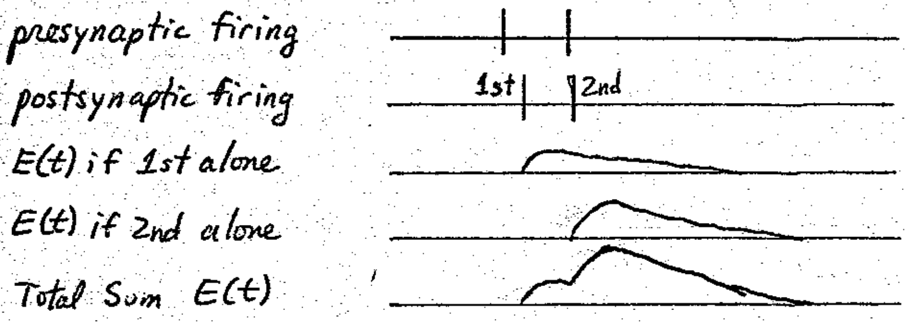


Figure 7. Illustration of Additive E(t).

number of firings just by adding up all their seperate E functions if alone.

The concept of adding up all the seperate E functions of each firing appears to be what is wanted theoretically, but do the difficulties in calculating this make it implausible? Fortunately, the equations can be chosen right so that exactly this concept can be implemented very simply. We chose E and an auxiliary function A this way:

9) $$\frac{dA}{dt} = -rA$$

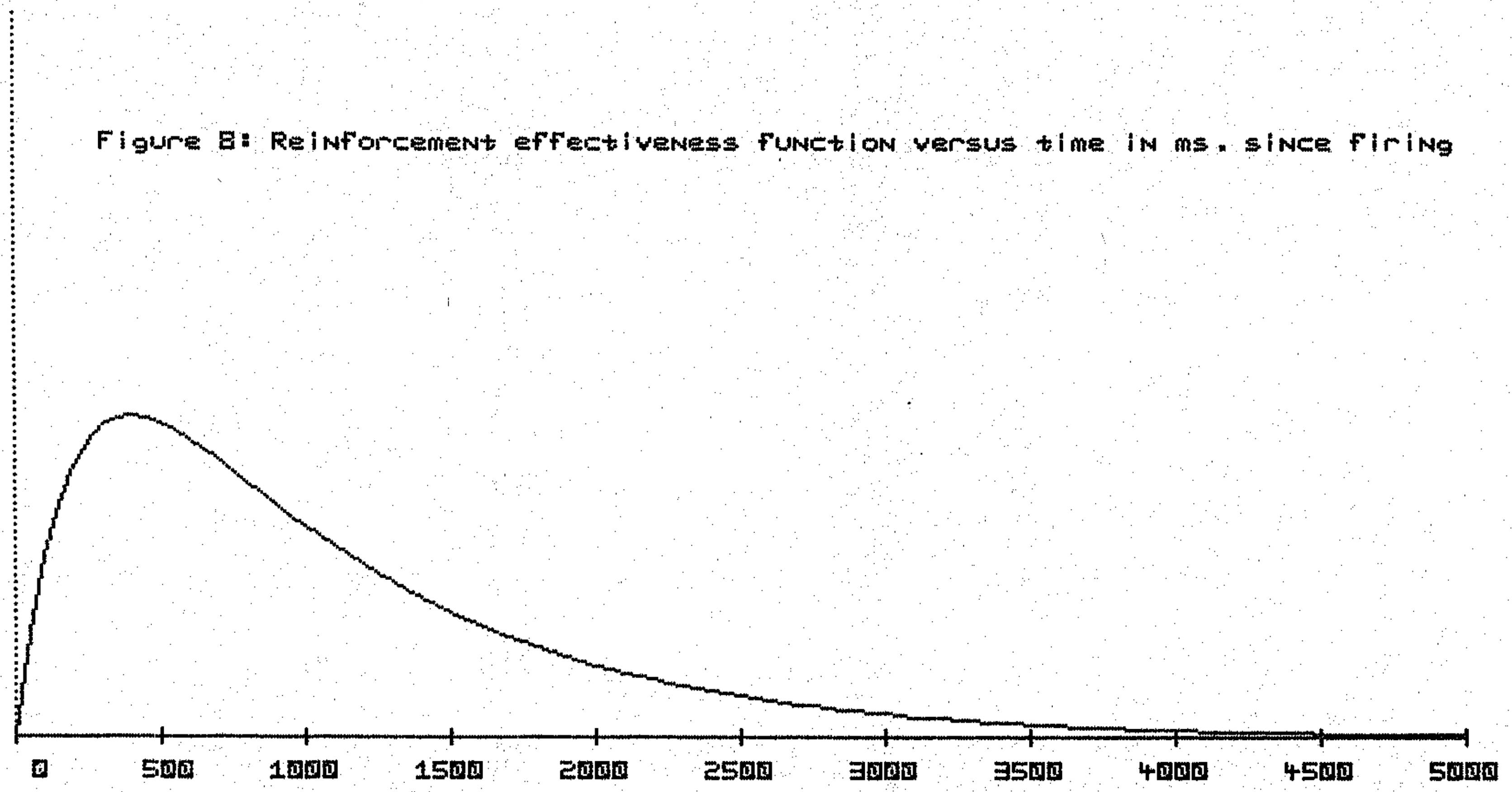
10) $$\frac{dE}{dt} = -sE + rA$$

Taking A(0) positive and E(0) zero, and solving the differential equations:

11) $E(t) \sim e^{-st} - e^{-rt}$ (see apendix A for derivation)

Which, for suitable r and s, is a very satisfactory reinforcement effectiveness function (see figure 8). These equations are used by adding C to A(t) each time there is a postsynaptic firing. A mathematical argument can then be used to show that the E function result of adding such a C to A is exactly the same as the E function if the C was not added plus the E function if the firing of effect C had occurred alone.

Figure 8: Reinforcement effectiveness function versus time in ms. since firing



It will be useful to note a few qualitative aspects of this eligibility function. First, a synapse's eligibility function is only significantly larger than zero if a signal has been passed by the synapse within the last few seconds. Second, its size is proportional to the contributions and number of such signal passing events. To a rough approximation it is just a measure of the extent of occurrence of such events, with a delay lag of 400 ms. Third, note that this function determines not how or in what direction the synapse changes, but merely how fast it changes.

4.2.2 Prediction Refinement -

Consider our synaptic change equation:

$$12) \quad \frac{dG}{dt}(t) \sim [R(t) - KG]\, E(t)\, Z(t-t1)$$

where t1 is time of last presynaptic firing

The term KG can be thought of as a prediction of what the reinforcement will be. If the reinforcement is precisely this much the synapse will not change. In the above equation the prediction is only a function of synaptic efficacy. This makes sense - other things being equal, the synaptic efficacy should be largest for those synapses that predict the most reinforcement. However, often other things are not equal. For instance, if a synapse if frequently presynaptically active, then even if it has a low synaptic strength it will have a large effect on the neuron. In general, we would like to make a synapse's effect on the neuron be the determinant of how much reinforcement it needs to remain unchanged. In this case, both synaptic strength and number and size of contributions should determine predicted reinforcement. This is apparently a multiplicative relationship – if the synapse contributes to two firings of the neuron then it should expect twice as much reinforcement. Thus I propose we need some prediction term P in our synaptic change equation thus:

13) $$\frac{dG}{dt}(t) \sim [R(t) - KPG]\, E(t)\, Z(t-t1)$$

I have tried to give a subjective justification for adding this P term. It has the added advantage of making the individual neuron a less unstable system, a problem that will be discussed later.

I will now try to state some of the qualitative properties of the P term so that can try to derive a quantitative formula for it. First, if there have been no firings for a while, and then there is one firing, then P should be the contribution of the synapse to the firing C (equation 4), as this times G is how much effect this synapse had on the firing. Second, if immediately after this firing we have other firings, then P should be approximately the sum of the synapse's contributions to the firings, because since the synapse is contributing to many firings, it is having a large effect on the neuron. This approximation is, of course, only accurate if the firings occur very close together, for as the get farther and farther apart they begin to look like seperate firings and we should apply the first case we just discussed where P is set to just the C of the most recent firing. Thirdly, if a firing occurs and the synapse did not contribute to it (C is zero), then P should not change, for the synapse has had neither more or less effect on the neuron. This last criterion points up that, interestingly enough, although P is not a constant, it is not particularly a funtion of time either. P's value is fixed except for when a firing to which the synapse contributes to occurs, at which point it makes a discrete jump to a new value.

Three qualitative criteria for how P should change in its discrete jumps have been given. To turn these into quantitative equations we will need to know a few of the mathematical properties of the E and A equations. These properties will be stated here and proved in appendix A.

Theorem: The area under the rest of an E funtion (integral of E(x) from t to infinity) is $\{E(t)+A(t)\} / s$.

Corollary: The area contributed to an E curve by adding a contribution C to A(t) is C/s.

Now the qualitative criteria can be stated more quantitativly:

1) if $A(t)+E(t) = 0$, then new $P = C$.
2) if $A(t)+E(t) =$ old P, then new P = old P + C.
3) if $C = 0$, then new P = old P.

Given these criteria I have settled on what seems to be the simplest suitable change of P equation:

$$14) \quad \text{new } P = \text{old } P \; \frac{A(t)+E(t)}{A(t)+E(t)+C\,[1-\{A(t)+E(t)\}/\text{old } P]} + C$$

I confess that even to me this is not a very intuitive formula, but taking cases it seems to have the desired properties. If C is zero, the complex fraction reduces to one and P is unchanged. If A(t)+E(t) = old P, as when there are two firings nearly together, the fraction is also one and P becomes old P + C. Finally, if there have been no firings for a while, so A(t)+E(t) is near zero (while C is not), then the fraction is near zero, and P = C.

With this added refinement of the P prediction term we can now ask about the value of the constant K in the synaptic change equation (equation 13). Consider a single isolated firing to which the synapse contributed an amount C. Call the highest value of its E function H. Suppose we want G to be stable when PG predicts the arrival of an amount of depolarization equivalent to the synapse's own depolarizing effect, if the predicted depolarization arrives at the optimal time (400 ms. after firing). Since the synapse is not to change when this occurs, then if the effect of zerosetting is ignored:

15) $$\int [R(t)-KPG]E(t)\,dt = 0$$

Also, excess reinforcement is equal to PG. To get excess reinforcement from weighted excess reinforcement we scale it so that reinforcement occurring at the optimal time has the unit effect:

16) $$\frac{\int R(t)E(t)\,dt}{H} = PG$$

so,

17) $\int R(t)E(t)\ dt = HPG$

H is defined to be the largest value of the E weighting function. Since the shape of the reinforcement effectiveness function is dependent only on the time constants (1/r and 1/s) and its height is determined by C for a single firing, we have:

18) $H = YC$

Where Y is a constant equal to the maximum value of E(t) when starting with original conditions $A(t0)=1, E(t0)=0$. In other words, this constant Y is well defined and dependent only on the constants r and s. From our theorem we know:

$$\int E(t)dt = [A(t0)+E(t0)]/s$$

Thus, for our single isolated firing (indicates $A(t0)=E(t0)=0$), since A gets incremented by C, we get:

$$C/s = \int E(t)dt$$

and,

$$C = s \int E(t)dt$$

so:

$$H = Ys \int E(t)dt \quad \text{from 9)}$$

so:

19) $\int R(t)E(t)dt = PGYs \int E(t)dt$ from 17)

and since from 15),

$$\int [R(t)-KPG]E(t)\ dt = 0$$

then:

$$\int R(t)E(t)dt = \int KPGE(t)dt$$

$$PGYs \int E(t)dt = KPG \int E(t)dt$$

(by 19 and since G can be arbitrarily constant)

so,

$$Ys = K$$

Thus our synaptic change equation is fully defined (except for an arbitrary rate of change constant) as:

20) $$\frac{dG}{dt}(t) \sim [R(t)-YsPG]\ E(t)\ Z(t-t1)$$

Where t1 is time of last presynaptic firing and with the constant approximately (zerosetting is being ignored) set so that the synapse will come to have as great an effect on the neuron as the effect is predicts to come in from another synapse. Since both inputs are by definition reinforcers, this means that with this constants, secondary reinforcement can become at maximum of equal effect as the reinforcement is predicts. There is no particular reason to give the constant this value, but this is a useful reference point in selecting a value for the constant.

4.3 Neuronal Element Stability

Neuron-like elements operating according to the equations given thus far are distinctly unstable. The instability is a basic consequence of using neural input as reinforcement to the neuron's synapses. The instability is the result of a positive feedback loop that is entirely local to the single neuron and its synapses, and would occur even if the input to the neuron remained constant in its characteristics. There are a number of ways to eliminate or minimize this instability that have interesting advantages and disadvantages, and most of which have not yet been extensively explored.

To show how the instability arrises, consider a neuron with many synapses. Assume the presynaptic neurons of these synapses fire in a totally random way with a fixed probability distribution. Also assume the initial average algabraic sum of inputs (reinforcement) to the neuron is zero (although it undergoes random fluxuations of course, depending on which of the presynaptic neurons happen to be firing and transmitting signals through their synapses. Consider what happens if the neuron fires and then, by chance, the reinforcement (input) following the firing happens to be slightly positive. Since this reinforcement is positive it will tend (in most cases) to make the synapses which caused the firing more positive if they were excitatory and less negative if they were inhibitory. In general, the positive reinforcement will result in changes to the synapses which will cause average input subsequently to be slightly higher than it was before, or in this case, slightly positive. Thus, when the neuron fires again it will probably get slightly positive input, which will cause new synapstic increases and thus further raise the level of average reinforcement. This process accelerates until it is completely irreversable and the neuron-like element is useless. A very similar positive feedback process occurs if the initial chance reinforcement is negative. In this case the synapses become smaller and smaller (more negative or less positive) to no useful purpose. The neuronal elements are generally unstable in that small fluxuations in their reinforcement are soon turned into large ones without any particular relation to environmental reinforcement dependencies.

One way to minimize the instability problem is to use for reinforcement not just how much input is being received now, but how much more is being received now than usually is. In other words, measure the reinforcement as input minus average input. This only minimizes the problem because the average can only be over some window size of past values and thus the average tends to trail behind what the real average has become because of learning changes in the synaptic efficacies. This solution can be made more exact by correcting the average immediately for any changes in the efficacies by claiming the neurons keep track of their presynaptic firing frequencies seperately. Thus, for a neuron with n synapses labelled 1 through n, it can use this equation to find average input from which to measure instantaneous reinforcement:

$$\overline{I} = \sum_{j=1}^{n} F_j G_j$$

where: $\overline{I}$ is average input
F is the frequency of presynaptic firing
G is the synaptic efficacy

This is a perfect solution to the instability problem but unfortunately is rather implausible for a biological neuron. In the computer simulations a variant of the straight average input was used to measure reinforcement from with satisfactory results.

4.4 Computer Simulations

To simulate the parallel operation of many neurons in a serial digital computer, time was divided into discrete intervals of about ten milliseconds. Although real neurons are not syncronized this way, it is a useful method of neural simulation which appears, in practice, to yield the same sort of results as more dificult continuous time simulations.

The crucial state variable of each neuron is its membrane potential. If a neuron's membrane potential exceeds its threshold the neuron is said to fire, and it provides input to other neurons. After firing, the membrane potential is reset to usually just less (more polarized) than resting potential. In the absence of input, the membrane potential decays exponentially to the resting level with a short time constant. A third source of change in membrane potential is input from other neurons. If there are n neurons, and Aij represents the efficacy of the synapse from neuron i to neuron j, and X is an n-element vector such that Xi =1 if neuron i is firing and zero otherwise, then this contribution to membrane potential can be written this way for a neuron i:

$$\sum_{m=1}^{n} A_{ij} X_{i}$$

The Aij's are allowed to be negative for inhibitory synapses and zero if there is no synapse between the two neurons. Changes in these synaptic efficacies was done according to the equations developed in the preceeding sections.

The basic learning process has been demonstrated as it was theorized to occur according to the two general descriptive principles. To show the learning process in full-fledged and many firings action and to illustrate at what level the neurons were simulated, an example will now be considered.

The modelled neural network is diagrammed in figure 9. The series of graphs labelled figure 10 show what happened in the network. The horizontal axis is time in hundreths of seconds. Each horizontal series shows the firing incidences in a single neuron. Here is a detailing of the network's connections:

```
neuron  1 is connected to no neurons
neuron  2 is connected to neurons 1, 5, 6, and 7
neuron  3          "              1, 5, 6, and 7
neuron  4          "              1, 5, 6, and 7
neuron  5          "              1, 6, and 7
neuron  6          "              1, 5, and 7
neuron  7          "              1, 5, and 6
```

Neurons 2, 3 and 4 receive no neural input and their firing is controlled by a simulated environment in the computer program. Neurons 1, 5, 6, and 7 do have synaptic connections from other neurons. Some of these synapses are initially nonzero and affect other neurons, but note that the synapses from neuron 4 seem to have near zero initial efficacy because there is little effect on the other neurons when it fires initially. Under certain conditions global positive reinforcement (positive input) is delivered to neurons 1, 5, 6, and 7. In this series of outputs,

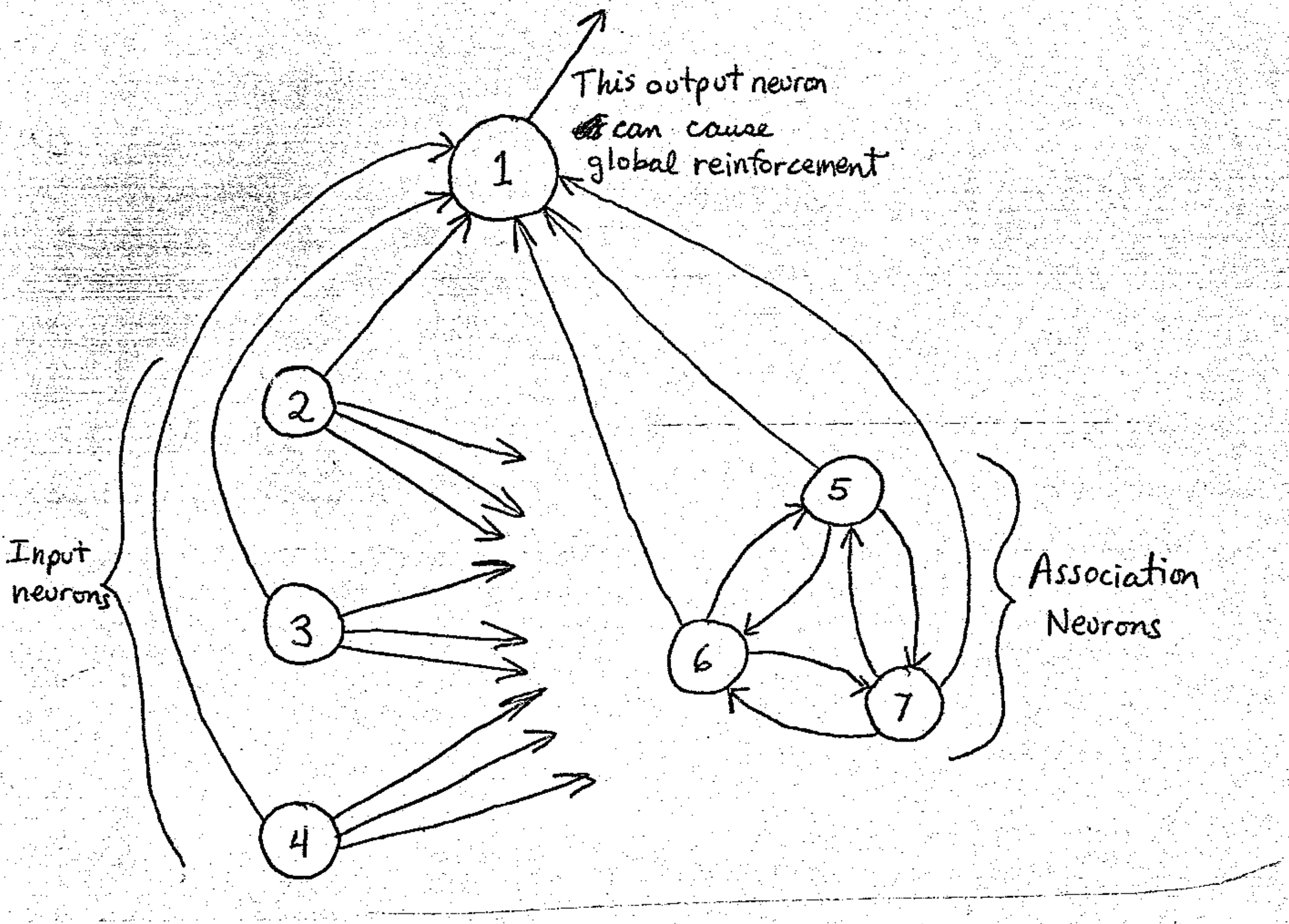


Figure 9. `A small network that was simulated. Result is in Figure 10.

global reinforcement is delivered if neuron 1 fires during a 200 ms. time interval which occurs at 7 second intervals. The presence of this interval is indicated by firing in neuron 4. The amount of reinforcement delivered in global reinforcement is a random variable with expected value proportional to the number of firings in this 200 ms. interval. The neural net setup is as in operant conditioning. Thus we find that neuron 1 learns to fire in response to neuron 4's firing. As this becomes learned, 4's firing becomes a predictor of reinforcement and neurons 5, 6, and 7 also learn to fire in response to 4's firing.

5.0 STABILITY OF LEARNING

A mind is an enormous organization. In my view, the principle reason it is able to have such organization is that it develops it gradually. Considering the great and careful organization of a mind it is all the more amazing that it can remain flexible and learn new things quickly. Somehow the brain can quickly take up new learned associations, gradually refine them into as abstract as necessary entities so that they are maximally accurate, and them remember them undisturbed by all the new things that it is quickly learning and gradually refining. I feel that it is necessary to provide at least some general explanation of how all this is possible. Such considerations will be paramount in further refining the equations the neurons should learn by. In the following subsections, I will discuss some of the characteristics I consider basic to a developing learning system as the brain appears to be, and I will show how the theory is particularly suited for systems that have such characteristics.

5.1 Neuronal Synaptic Competition

In the computer simulations, neuronal reinforcement was taken as the algebraic sum of the inputs measured from its weighted average. The average was weighted by E(t), the reinforcement affectiveness. The result is that neuron compares reinforcement it receives after a firing not to just the reinforcement it usually receives, but to the reinforcement it usually receives after a firing. The intention here is to develop competition among the synapses of the neuron. Using this measure of reinforcement, only those synapses of strongest positive reinforcement prediction or contingency indication would become positive. This can prevent less valuable learning from interfering with already present important learning.

5.2 Symbol Distribution

The model neurons are the unit learning elements of the single channel theory. They decide which of their inputs to listen to and in this way eventually determine the animals responses. In order for crucial neurons to choose to fire to a particular stimulus, some signal affected by that stimulus must be accessable to them as one of their synaptic inputs. It is vital to get the important symbols for decision making to the relevent neurons. This is all the more important to do in the single channel theory because the neurons will spend all their efforts figuring out how to manipulate their own inputs, as these are their reinforcement. In order for this to be useful to the whole system there should be some control over what these inputs are. One way to do this is to increase the range of distribution of signals that can be shown to be related to reinforcement. This happens naturally in the single channel theory because, as a secondary reinforcer, an input stretches over the range that it predicts reinforcement for. Thinking of a range and its reinforcement, such as the whole brain and global reinforcement, we see that the neurons of the brain have access to signals according to how related the signals are to the neurons' receiving reinforcement from other sources.

6.0 REQUIREMENTS OF A SINGLE CHANNEL NETWORK

6.1 Matched Competition

In a system in which only one kind of signal is used for both reinforcement and information transfer, one has to be careful of the times when these two functions interfere with each other. Here we will discuss one such problem and its solution by proper network construction. Although the reinforcement to a neuron is excitation, when the system delivers reinforcement for a response, it does not particularly *want* the response to be repeated again. In principle, this problem is easily solvable. Clearly the reinforcement does not go particularly to the neurons of the response, but to all the neurons the response neurons were chosen from (the response neurons were just the only ones to learn to the reinforcement). Since the reinforcement will go to all those neurons to be chosen amongst, with none particularly favored, the particular response will not tend to be repeated. However, there remains some problem in that there is all this extra excitation (or inhibition, as the case may be) coming in to the neurons on reinforcement - what kind of effects might it have?

I feel that the wisest course here is to propose that the brain is so constructed so as to minimize the effects of general nonspecific reinforcement or excitation. Probably neural nets consisting of both excitatory and inhibitory connections can be made so that nonspecific reinforcement only changes the firing rate of the neurons involved, but not the overall excitatory or inhibitory nature of their output. Now we need only explain why there is little external behavior change upon global reinforcement deliverance and increased firing rates of the neurons. One solution is to propose that the outermost parts of the output system that do not need to learn (perhaps the motoneurons) do not receive reinforcement, but only the output from the network that receives reinforcement. Since we have said the relative composition of this output is unchanged, there is little necessary external response of the organism to global reinforcement occurrence. This necessary arrangement of in some way equivalent excitatory and inhibitory connections from the neurons receiving reinforcement to those controlling the muscles so that general activity changes in the former dont affect the latter is called matched competition.

6.2 Zerosetting Experimentally Disproved?

The proposed zerosetting mechanism prevents reinforcement from being effective in changing a synaptic efficacy if that synapse has very recently been presynaptically active. This is clearly necessary or else any long duration stimulus (lasting over 400 ms. for instance) will start to act as its own reinforcement for getting the neuron to fire in response to it a little while ago. As we discussed earlier, zerosetting is necessary in general to force the reinforcement to come in through feedback loops containing the environment. However, despite all these reasons why we need zerosetting for the single channel theory, learning experiments have not shown evidence of such a mechanism.

In classical conditioning, the UCS is the reinforcer and the CS is the input signal that the neurons responsible for the UCR should learn to. It would appear that zerosetting would prevent classical conditioning if the CS and the UCS totally overlapped in time as diagramed in figure 11. However, it is found experimentally that the crucial variable that determines learning is the time interval between CS onset and UCS onset. There seems to be little diference between the delayed procedure in which the CS starts before the UCS but may overlap it partially or totally, and the trace procedure in which the CS starts and ends before the UCS onset (Tarpy, p 19-20). However, although this result is by no means support for the single channel theory, it is also not totally inconsistent with it.

The experimental result can be explained by claiming that the apparent contradiction with the theory arose because of our mistaken assumption of identity between the external CS and the internal neural signals that signify it to some extent and are what is actually involved in the learning. In a general way, the nervous system is known to be more sensitive to changes in stimulation than to the absolute magnitude of the stimulation. Thus, it is not unreasonable to propose that the nervous system is more sensitive to the onset of the external CS than to whether it remains on or not. Specifically, there are probably many neurons that respond to the onset of the external CS but which do

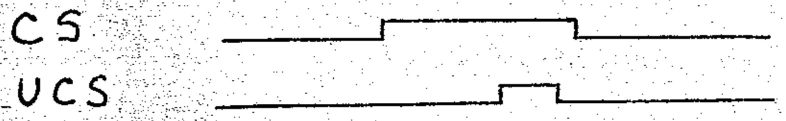


Figure 11. A CS that totally overlaps its UCS on reinforcement,

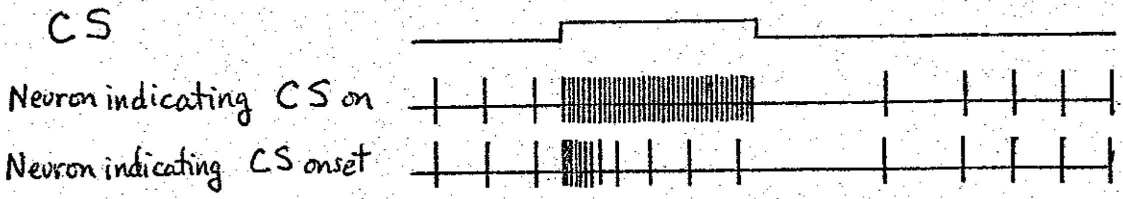


Figure 12. Illustration of stimulus on and stimulus onset neuron behavior,

not indicate by their firing whether or not the external CS continues. These neurons would thus constitute signals indicating CS onset rather than CS presence. The kind of difference in firing patterns is shown in figure 12.

Presumably stimulus onset neurons arise because of some form of of neural inhibition which is known to pervade the central nervous system. If stimulus onset signals are actually the dominant mode of symbol in the nervous system then zerosetting would not be effective internally in preventing learning, even if the external stimulus continues while the UCS as reinforcement occurs.

6.3 The Expected Level Of Reinforcement Problem

There is experimental evidence from learning experiments that brains develop a situation dependent expection of a level of reinforcement. The simplest example is avoidance conditioning. If an experimenter repeatedly follows tone presentation by electric shock to a rat, then the rat will learn to expect the negative reinforcement level after the tone. This is shown by allowing the rat to perform some response such as bar pressing to prevent the arrival of the shock, which the rat will learn to do. The theoretical problem here is that the rat received no apparent positive reinforcement for pressing the bar, all it received was the absence of negative reinforcement. So why did it learn to perform the response? At present I do not know how to explain this with the single channel theory. I can think of three possible kinds of explanations, all of which may be partially true:

1) There is no real expected level of reinforcement involved here. The bar pressing response is something the rat may or may not do. At some neural level these two choices are of equivalent status. One is not the absence of the other, they are just the two sides of a choice. Thus, when the rat has access to the bar and chooses not to press it, this gets punished, so the rat learns in the usual way not to not press th bar. Thus, he learns to make the other choice, bar pressing, by elimination. This seems like it might not be an effective method of learning, but it is well known that punishment avoidance conditioning is frequently much less effective than conditioning using positive reinforcement.

2) The level of reinforcement expectation ability of the brain somehow emerges as a network phenomena that is not present in the individual neuron.

3) The mathematical formalization of the neuronal element that I have presented in this paper is not yet right, but must be modified to give it this ability. At present, this seems the most promising idea. The equations can easily be changed in ways that apparently would make the neurons measure reinforcement from an expexted level in changing the synaptic efficacies. Only preliminary computer simulation experiments have been tried with this idea and the results are unclear at present.

7.0 CONCLUSION

The basic concepts behind the single channel theory have been around publicly since 1972 in a wide-ranging report by Klopf. For one reason or another, the theory has not become well known. In this paper I have tried to establish the importance of this theory by limiting myself to making two main points. First, the theory has great potential for explaining the perplexing results of learning theory experiments at a mechanism level. The theory provides an interesting common process explanation of operant and classical conditioning, and as a natural extension, the theory will also explain secondary reinforcement, including chaining and the mechanism behind secondary reinforcement cues. Much of the theory's potential comes from being a local reinforcement theory. A local reinforcement theory appears to be necessary to explain sensory preconditioning and latent learning. The single channel theory explains sensory preconditioning as classical conditioning of internal responses. The theory looks like it will be able to explain latent learning also, as an analogous extention of operant conditioning, but little attempt has been made at this point to give the mathematical models the full ability for latent learning. The theory is also not yet able to fully explain avoidance conditioning, but neither this nor latent learning appear irreconcilable with the single channel theory. The theory is still young and simple, and I expect progress to be made towards explaining these phenomena. For now I want to point up the theory's abilities in providing mechanism level explanations (rather than just descriptions) of the fundamental processes of learning as revealed experimentally.

My second main point about the single channel theory is that it is plausible. There was some question about the theory's constructability. The mathematical formalization and computer simulations reported in this paper have now made it clear that the theory has no unresolvable mathematical or logical problems. A more dificult plausibility question is whether it is plausible that a suitably arranged network of neurons operating according to the theory will organize themselves as complexly, efficiently, and as flexibly as brains are known to do. I tried to show this kind of plausibility in the last sections of this paper by discussing stability of learning, network arrangement, and symbol construction and distribution.

REFERENCES

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

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

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

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

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

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

APPENDIX A

E FUNCTION AREA THEOREM

First we will solve the differential equations to find a general equation for the E function.

given:

1) $$\frac{dA}{dt} = -rA$$

and

2) $$\frac{dE}{dt} = -sE + rA$$

find E(t) and $\int E(t)dt$ from zero to infinity.

3) $$A(t) = A(0)\,e^{-rt} \qquad \text{from 1)}$$

4) $$sE + \frac{dE}{dt} = rA(0)\,e^{-rt} \qquad \text{from 2), 3)}$$

This is a first order linear differential equation in E, and it has the following textbook solution:

5) $$E(t) = \frac{-rA(0)}{r - s}\,e^{-rt} + c\,e^{-st}$$

Now find c in terms of E(0):

$$E(0) = \frac{-rA(t)}{r - s} + c \qquad \text{from 5)}$$

6) $$c = E(t) - \frac{-rA(0)}{r - s}$$

so,

7) $$E(t) = \frac{-rA(0)}{r - s}\{e^{-rt} - e^{-st}\} + E(0)\,e^{-st} \qquad \text{from 5), 6)}$$

8) $$\int_0^{inf} E(t)dt = \frac{-rA(0)}{r - s}\{\frac{-1}{r}e^{-rt} - \frac{-1}{s}e^{-st}\} - \frac{1}{s}E(0)\,e^{-st}\Big]_0^{inf}$$

$$= \frac{-rA(0)}{r - s}\{1/s - 1/r\} - E(0)/s$$

$$= A(0)/s + E(0)/s = \{A(0)+E(0)\}/s$$

Theoram: The area under the rest of an E funtion (integral of E(x) from t to infinity) is {E(t)+A(t)} / s.

Corollary: The area contributed to an E curve by adding a contribution C to A(t) is C/s.