A UNIFIED THEORY OF EXPECTATION IN CLASSICAL AND INSTRUMENTAL CONDITIONING

Richard S. Sutton

A UNIFIED THEORY OF EXPECTATION IN CLASSICAL AND INSTRUMENTAL CONDITIONING

Richard S. Sutton

The study of animal learning has been an important, even central part of psychology in the past. The ability to learn can be seen as the mind's most salient characteristic. Certainly a human mind is shaped to an enormous degree by experiential as well as genetic factors. The adult human mind is very complex, but the question remains open whether the learning processes that constructed it in interaction with the environment are similarly complex. Much evidence and many peoples' intuitions suggest that the learning processes are in fact simple and that the adult mind's complexity is due to a long history of adaptive interaction with a complex environment. Among such suggestive evidence is an impression emerging from the study of human memory development that the basic memorial "hardware" processes, such as forgetting and inter-memory transfer, do not improve with age, and that other factors must be responsible for developmental performance improvements (Olson, 1976). Another source of evidence indicating simple learning processes relative to the mind's complexity is the experimental work on learning showing that the same sort of learning processes appear to be operating in animals ranging in complexity from pigeon to man.

The investigator of animal learning usually works on the premise that there are general laws of learning that apply to some extent to humans as well as to his animal subjects. Faced with the enormous complexity of any animal's behavior, a natural goal for the theorist of animal learning is to formulate a sufficient theory of learning that is not only general, but also as simple as possible. One of the major results of this approach has been the development of the terminology of stimulus and response. The behavior of all animals can be formulated in stimulus-response (S-R) terms, and this language appears sufficient to describe many learning regularities across species. As the psychology of animal learning progressed, one of the theorists' major frustrations became the fact that there were two different ways of training an animal that bore several similarities, and yet which he could not reduce to a single learning process. The two different ways of training are the well known classical, or Pavlovian, and instrumental, or Thorndikian, conditioning. The major point of this paper will be to present a unified theory of these two forms of conditioning. Rescorla and Solomon do an excellent job summarizing the essence of the difference between the two training paradigms, and also point up the importance that has been given to resolving the relationship between them:

> "The procedures which the experimenter (E) carries out in the Pavlovian, or classical, conditioning experiment are quite different from

those he carries out in a Thorndikian, or instrumental, training experiment. In the Pavlovian conditioning experiment, E ideally has full control over all experimental events; he determines the time of occurrence and the duration of a trial without any regard to the animal's behavior. That is, E arranges relations between stimulus events which he controls. In contrast, in the Thorndikian training experiment, E only arranges it such that the animal's behavior at specified times will yield predetermined environmental changes; E arranges relations between the animal's behavior and future stimulus events.

"Because the laws of learning are stated as interrelationships between experimental operations and consequent behavioral changes, the laws of [Pavlovian] conditioning and those of [Thorndikian] learning must be different at a descriptive level. This would be so even though the behavioral changes were identical in the two cases. On the other hand, if some theoretical system could be developed to unify the different empirical laws, to reduce them to the same general underlying principles, then the laws of behavioral modification would deduce the outcomes of both types of experiment. In an important sense, the history of learning theories is a succession of attempts to specify the relation between the outcomes of the two types of experiment." (Rescorla & Solomon, 1967, p. 151-152)

Like most other single process learning theories, the resolution of the relationship between classical and instrumental conditioning presented in this paper will propose that the outcomes of the two different experimental procedures are the result of a common underlying learning process. This learning process derives from a startlingly simple idea found by A. Harry Klopf while considering similarities between neural and social goal-seeking systems (Klopf, 1972). As I see it, the major insight that culminated in Klopf's idea was that a single kind of neuronal signal could be used both to evaluate and to perform neuronal actions. This insight can also be traced back to previous work (V. Griffith, 1962, 1963; Wilkins, 1970), but Klopf was the first to identify neuronal reinforcement with neuronal input, excitation as positive, and inhibition as negative. The theory presented here has evolved considerably but directly from Klopf's. Previous work dealing with Klopf's idea has treated it primarily as a theory of the neuron. This paper will develop a theory of learning using the traditional language of stimulus and response based on his idea. The important neuronal aspects of Klopf's idea will be treated primarily as a possible mechanism behind the abstract learning theory.

In the first part of this paper a single process learning theory will be presented and applied to the most straightforward classical and instrumental conditioning paradigms. The second part of this paper will go considerably beyond this to consider what can be collectively called "expectation phenomena" which occur in both classical and instrumental conditioning and which require greater theoretical sophistication for explanation. Two major classes of expectation phenomena are considered: a) phenomena involving expectations that influence the effectiveness of reinforcement, such as avoidance conditioning, and b) phenomena where some stimuli <u>dominate</u> others in that only they, or primarily they, are associated with the response at the expense of the others, such as overshadowing and blocking. This paper will show that both classes of expectation phenomena, in their realizations in both classical and instrumental conditioning, can be derived from the postulation of a single process involving an expectation term. Such a unified explanation of major expectation phenomena occurring in both conditioning paradigms constitutes strong evidence for this learning process as a general common process explanation of classical and instrumental conditioning.

PART I : THE SINGLE PROCESS THEORY

Any theory of learning contains, either explicitly or implicitly, a model of the mind, whether consisting of associations, stimuli, and responses or of information stores, channels, and processors. This is the framework within which rules for the dynamic operation of the system are stated. The framework or model of the mind for the learning theory being developed in this paper will be presented in the following section. The learning process that operates within and on this framework will be presented in a second section. A third section will describe the theory's explanations of some simple learning phenomena and a fourth will discuss the neuronal basis of the theory. Finally, a fifth section will review the major constructs of the theory and introduce a useful parallel terminology.

The Framework or Mind Model

1. The mind's state is determined by the activation levels of all of a large number of modes.

2. External stimuli have their presence reflected in the activation of corresponding modes. These will be called stimulus reflecting modes, or simply stimulus modes.

3. External responses occur when the corresponding mode or modes become sufficiently activated relative to antagonistic <u>response causing modes</u>.

4. The change in level of activation of each mode depends on the activation levels of the other modes as well as the current configuration of external stimuli. As a mode's level of activation increases, it may have a greater effect on each of the other modes' levels of activation. The effect of the increased activation of a particular mode on another is determined by the <u>association</u> from the first mode to the second.

5. The learning process has its effect on the associations between modes (i. e. on the modes influence of each other's level of activation). Thus it determines the succession of modes and responses, dependent on external input.

Although some modes are being called stimulus reflecting modes, and some response causing modes, these modes are in no other respects different from modes in general. In addition, a single mode may be both stimulus reflecting and response causing. Such a mode would result in an unlearned stimulus-response reflex. The usual case, however, is that at least one learned association will have to be utilized for a stimulus reflecting mode to affect behavior.

The Learning Process

At this point it will be useful to develop a symbolic notation.

> Notation: Let A_j ($A_j(t)$) denote the level of activation of mode j (at time t).
>
> Notation: Let iV_j ($iV_j(t)$) denote the sign and magnitude of the association from mode i to mode j (at time t). A positive association indicates that high activation of mode i tends to increase the activation of mode j, and a negative association indicates that high activation of mode i causes lowered activation of mode j. The larger the absolute value of the association the larger is the effect of mode i on mode j. In other words, the change in the activation of mode j because of the level of activation of mode i would be related to $Ai \cdot iV_j$.

All modes are said to be continuously generating an <u>expectation</u> of what the level of activation will be. The current expected level of activation (the expectation) of a mode is based on the recent activation level of the mode. The higher the level of activation within the last few seconds, the higher the level expected for the present, and the lower the recent activation level, the lower the expectation for the present. It will be sufficient to treat expectation of level of activation at a certain time as approximately an average of recent prior activation levels.

> Notation: Let P_j ($P_j(t)$) denote the expected level of activation of mode j (at time t). To a first approximation $P_j(t)$ is proportional to average$[A_j(t-T)]$, the average level of activation of this same mode some small time interval T (a few seconds or less) previous. (The letter P is used because expectation of level of activation can also be thought of as a <u>prediction</u> of level of

activation)

As has been stated previously, it is the association between modes, the $_iV_j$, that are changed by the learning process. The learning process only changes an association $_iV_j$ to the extent that modes i and j have been simultaneously active in the recent past. The concept of recent past is treated similarily to the way it was for expectation.

> Notation: Let $_iE_j$ ($_iE_j(t)$) denote the eligibility of the association $_iV_j$ for undergoing changes as per the learning process (at time t). To a first approximation $_iE_j(t)$ can be said to be proportional to average[Ai(t-T)A_j(t-T)], the average product, or logical AND, of the two modes' activation levels some small time interval T previous.

Now the learning process can be stated: an association $_iV_j$ between two modes, i and j, is eligible for change to the extent that modes i and j have been simultaneously active within the last few seconds. To the extent that an association is eligible, it will change according to the difference between the activation A_j and the expected activation P_j of the affected (destination) mode. If the activation level is greater than the expected level, then the association is increased, but if the expected level exceeds the actual activation level, then the association is decreased. In symbols:

$$\text{i)} \qquad \frac{d}{dt}\, iV_j = iC_j\,(\,A_j - P_j\,)\; iE_j$$

where: iC_j is a constant depending on the particular association being changed

The learning process is assumed to be constantly acting on all associations between all ordered pairs of modes. The direct effect of the environment on stimulus modes and of response modes on the environment is assumed not to be subject to this learning process, but to be permanently fixed.

Simple Learning Phenomena

The learning process will now be illustrated by using it to explain a few of the most basic learning phenomena: classical conditioning, instrumental conditioning, and secondary reinforcement.

Classical Conditioning

In a simple classical conditioning experiment the subject is repeatedly presented with a neutral conditioned stimulus (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, the conditioned response (CR), which closely resembles the UCR or some part of it. In a typical classical conditioning experiment, a dog is repeatedly presented with first the sound of a bell (the CS), and then its food (the UCS), which causes the dog to salivate (the UCR). Eventually, the sound of the bell alone causes salivation (the CR).

The mind model and the learning process provide an explanation of this basic classical conditioning experiment as follows. The CS and the UCS both have their occurrence reflected in heightened activation of corresponding stimulus modes. Since the UCS reflexively causes the UCR, it can be considered given that the mode excited by the UCS also causes the UCR when it is sufficiently activated. The training sequence is first increased activation of the mode excited by the CS, followed by heightened activation of the mode excited by the UCS and which causes the UCR. The heightened activation of the CS excited mode will result in increases in eligibility of the associations going to and from this mode. Consider the associations from this mode to the other modes. The eligibilities for these associations are high during the short time after the CS when the UCS occurs and increases activation of its mode. Since this mode experiences an increase in activation, at this moment its expectation of activation is still low while its activation is high. According to the learning process this will result in increases in the associations to the mode in proportion to how eligible they are. Since the association from the CS excited mode to this, the the UCS excited and UCR causing causing mode, will have a high eligibility at this time, this association will be increased. The result is that now the occurrence of the CS will be more likely to result in the emission of the UCR because now activation of the CS excited mode tends to excite, or increase the activation level of, the UCR causing mode.

Once this process has begun, once the CS reflecting mode starts to excite the UCR causing mode, the eligibility will start to become larger on subsequent trials, as both modes will be unusually active as the CS occurs, and the learning process will be accelerated. This acceleration will reverse, and the size of the association will stabilize, as the association approaches that size at which an occurrence of the CS causes as great an activation level of the UCR causing mode as is subsequently caused by the occurrence of the UCS. No further learning occurs in this situation because the activation of the UCR mode caused by the CS generates a subsequent expectation of activation which just matches the activation caused by the UCS. When activation equals expectation, the hypothesized learning process dictates that no further learning changes will occur.

Instrumental Conditioning

In instrumental conditioning a stimulus event called an instrumental reinforcer is presented to the subject contingent on his performance of a selected conditioned response (CR). Frequently, the instrumental reinforcer is presented only if the CR occurs in the presence of some conditioned stimulus (CS). The result is a change in the probability of occurrence of the CR in response to the CS. The instrumental reinforcer is called a positive instrumental reinforcer if the result is that the CS is more likely to lead to the CR, and a negative instrumental reinforcer if the result is that the CS is less likely to lead to the CR. In a typical instrumental conditioning experiment, a rat is rewarded with food (the positive instrumental reinforcer) whenever he presses a bar (the CR) in the presence of an auditory tone (the CS), and never when the tone is absent. Eventually the rat will learn to press the bar when, and only when, the tone is present.

The mind model and the learning process provide an explanation of this instrumental conditioning paradigm as follows. The CS, as in classical conditioning, has its occurrence reflected in the activation of a stimulus mode. The instrumental (assume positive) reinforcer is a stimulus that excites a high proportion of the mind's modes. Heightened activation of a certain response mode relative to its competitors for expression will result in the emission of the CR. Consider the training sequence from the point of view of the mode causing the CR, with particular attention to its association from the CS reflecting mode. Occasionally the CR causing mode becomes active enough to be expressed – for the CR to occur. If the CS is not present when this happens, then the eligibility of the association from the CS reflecting mode to the CR causing mode will be small relative to what it would be if the CS was occurring when the CR was emitted. If the CS is not present when the CR is performed (eligibility low), then the instrumental reinforcer does not occur, but since the eligibility is low, little change is made in the strength of the association from the CS reflecting to the CR causing mode. If the CS is present when the CR is performed (eligibility high), then the instrumental reinforcer occurs, causing a moment of slightly higher activation than expectation for the CR causing mode following its causing of the CR. Since the eligibility is high at this time the association from the CS reflecting mode is increased. Next time the CS occurs the mode reflecting it will excite the mode that causes the CR, and make the CR more likely to occur.

As in classical conditioning, this is initially an accelerating process – the learning is strengthened each time the CR occurs to the CS, and the result of the learning is to make this sequence more likely. The increase in the association from the CS mode to the CR mode will be halted when an occurrence of the CS causes as much activation of the CR mode as that caused by the instrumental reinforcer.

Secondary Reinforcement

The mind model and the learning process also provide an interesting and simple explanation of an important learning phenomena that has not yet been mentioned in this paper: secondary reinforcement. Almost all experiments are believed to involve secondary reinforcement, although usually not explicitly. For example, when an animal is being rewarded with food being deposited in a hamper, it soons starts to act as if it were being rewarded not by the eating of the food, but by the sounds and other incidental stimuli that first indicate that the food is going to be delivered. Stimuli which are correlated with the subsequent occurrence of an instrumental reinforcer come to have similar reinforcing capabilities. They are said to have become secondary reinforcers.

The development of secondary reinforcers is consistent with the proposed mind model and learning process. The instrumental reinforcer, as in the explanation of instrumental conditioning, is assumed to cause global excitation (or depression) of a high proportion of the modes. If the global reinforcement is positive, then the activation levels of the modes are increased, while their preceding level of activation, and thus their expectation, remains momentarily at the former lower level. Thus, the associations to each mode will all tend to be increased, and among them the tendency will be in proportion to their eligibility (for activation and expectation are common to all associations to a mode). The eligibilities of the associations to a mode will be distributed in proportion to the average recent activation levels of the modes that are the sources of the associations, as the average recent activation of the destination mode is also constant among them, and since eligibility is the average product of these two activations. Thus, those associations that lead from modes that were highly active previous to the occurrence of the global, or instrumental, reinforcer will all be increased. Similarly, it follows that if the instrumental reinforcer had been negative, then the activation would have been less than expectation, and the associations from previously highly active modes would have been decreased. The result is that the modes whose high activation preceded the instrumental reinforcer will have an effect on the other modes more like that had by the instrumental reinforcer. If a stimulus repeatedly precedes an instrumental reinforcer, its effect on the other modes will continue to change until it is equal to that of the instrumental reinforcer.

The Neuronal Basis of the Theory

As mentioned in the introduction, the present theory evolved from a theory by A. Harry Klopf (Klopf, 1972). Klopf's theory of the brain's operation was stated in terms of neurons, and the present theory, though thus far stated in a more abstract, learning-theoretic manner, should retain this neuronal history. An attempt will now be made to translate the current theory into neuronal terms. This can be viewed both as a possible mechanism for the theory, and as a further specification and development of the constructs of the theory. The neuronal interpretation should not be treated as no more than a mechanism, for it does have behavioral implications, and could be used to explain certain phenomena, although it will not be used in this way in this paper.

The neuronal interpretation of the theory has not been used as its primary statement for two reasons. First, such a mechanistic and physiological theory is unfamiliar to cognitive psychology. Second, such a neuronal statement is at a level of detail that would require great specification of assumptions, both neurophysiological and functional. By giving the neuronal interpretation status primarily as a mechanism allowance is made for many of these neuronal assumptions to be wrong. If neurons are found not to behave exactly as postulated, the mechanism may need to be rejected, but the behavioral learning theory can remain relatively intact.

The basic tenants of a neuronal interpretation of the mind model and learning process previously described are as follows:

1. Each mode consists of a population of neurons. A mode's level of activation is the average rate of firing or impulse generation of the mode's neurons. A single neuron can participate in several modes.

2. The associations between modes are the synaptic connections from the neurons of one mode to the neurons of the other mode. A significantly positive association between two modes i and j (${}_iV_j \gg 0$) means the synapses connecting presynaptic neurons of mode i and postsynaptic neurons of mode j are predominantly excitatory and transmission efficient. The constant ${}_iC_j$ in the modal learning process represents the extent to which the neurons of mode i have synapses to the neurons of mode j. More salient stimuli can be expected to influence more neurons and thus to have larger constants associated with their associations.

3. The efficiency of synapses in transmitting impulses from neuron to neuron is what is proposed to be changed by the learning process. To represent the neuronal learning process, we will need a neuronal symbolism. The following

neuronal symbolism is directly analogous to the modal symbolism and thus should require little additional effort to understand. However, as the other sections of the paper will make no reference to the following neuronal symbolism, the rest of this section may be safely skipped by those not interested in a neuronal formalization of the theory.

> Notation: Let iW_j ($iW_j(t)$) represent the efficiency of the synapse from neuron i to neuron j (at time t). A negative value for iW_j indicates an inhibitory synapse. (iW_j corresponds to iV_j, the association between modes i and j)
>
> Notation: Let X_j ($X_j(t)$) represent some measure of the current firing rate of neuron j (at time t). (Corresponding to A_j, the activation level of mode j) The simplest measure this could be is an instantaneous measure: $X_j = 1$ when neuron j is firing, and $X_j = 0$ otherwise.
>
> Notation: Let $P_j'(t)$ be a measure of the average rate of firing of neuron j a short time previous. (Corresponding to $P_j(t)$, the average recent activity of mode j)
>
> Notation: Let $iE_j'(t)$ be a measure of how much neuron i fired such as to have the effect of its firing influence firings of neuron j, averaged over a short time before t, the present. This is a measure of how much the synapse from neuron i to neuron j would have affected neuron j's recent firing activity, if the synapse had been efficient. (This corresponds to $iE_j(t)$, the eligibility of the association between modes i and j)

Each synapse of each neuron is hypothesized to carry out the learning process:

$$2) \qquad \frac{d}{dt}\, iW_j \text{ "=" } (\, X_j - P_j' \,)\; iE_j'$$

The equal sign has been quoted because, based on neurophysiological evidence, neurons appear to have their synapses fixed to be either permanently excitatory or permanently inhibitory. The action indicated by the above learning equation is not carried out if it would take the synapse beyond its allowable boundaries: excitatory synapses cannot become inhibitory (negative), and inhibitory synapses cannot become excitatory (positive), and synapses can become neither infinitely large nor infinitely small without bound - instead they saturate at some finite value.

Reinforcement, Activation, Expectation, and Effective Reinforcement

An important thrust of the theory presented in this paper, derivative from Klopf's work, is that at the level of modes, activation level acts as reinforcement in causing changes in associations between modes. In the discussion of instrumental conditioning it was hypothesized that instrumental reinforcers are stimuli that affect the level of activation of all or almost all modes. In the discussion of classical conditioning it was the arrival of the UCS which affected the level of activation of the CR causing mode, and which was necessary for learning changes to occur. In classical conditioning the UCS is often spoken of as the reinforcer for the S-R connection between the CS and the CR. In both kinds of conditioning, it is reasonable to talk of modal level of activation as reinforcement, with high activation acting as positive reinforcement and low activation acting as negative reinforcement.

A careful reading of the hypothesized learning process reveals both the utility and the folly of speaking of activation as reinforcement. According to the learning process, it is the destination mode's activation level minus expectation level that determines whether the association is increased or decreased - whether the CS increases or decreases in its ability to cause the CR. Recall that expectation, or expected level of activation, can be approximated by an average measure of the recent activation level. Thus, if it can be assumed that recent activation has been at normal levels, then high activation does act to increase the association (as positive reinforcement), and low activation does act to decrease the association (as negative reinforcement). Thus it is reasonable in many cases to speak of level of activation as modal reinforcement. However, in many other cases, recent activation, and thus current expectation, will not be at normal levels, and this way of speaking will not be valid. For instance, if expectation is much higher than usual, then a slightly high level of activation would be effective as if it was negative reinforcement, decreasing the association.

A new parallel terminology will be used henceforth to clarify and capture these ideas. Activation level will also be called reinforcement level. Expectation of activation level will also be called expectation of reinforcement level. However, the difference between these two, that which actually determines whether the associations to the mode are increased or decreased, will be called <u>effective reinforcement</u>. An interesting point of view that arises out of this terminology is that the effective reinforcement, that which determines learning changes, is said to be given by the difference between expected reinforcement and received reinforcement.

It is important to realize that throughout the previous discussion modal rather than global reinforcement was being discussed. The theory's use of only modal reinforcement, a local reinforcement, adds a new dimension of complexity to applications of the theory to learning experiments. With different reinforcements, different evaluations of performance, simultaneously active in different modes, learning changes are likely to not be uniform throughout the mind, but to vary from mode to mode, depending on their local modal reinforcements. Global reinforcement can only be discussed to the extent that a significantly large proportion of the modes are undergoing similar changes in level of activation.

PART II: EXPECTATION PHENOMENA IN CLASSICAL AND INSTRUMENTAL CONDITIONING

The essential concept behind the single process explanation of classical and instrumental conditioning as two aspects of a single process was provided by Klopf (Klopf, 1972). The major extension of Klopf's attempt's to explain learning phenomena presented herein is the use of an expectation factor as an important determinant of learning changes. Since the learning process uses an expectation factor, and since the ~~learning~~ learning process is used to explain both classical and instrumental conditioning, there should be behavioral implications of the expectation factor for both kinds of conditioning. As evidence for a unified theory of expectation, the second part of this paper works out two such implications and verifies them against the known experimental results. The two behavioral implications of the theory for both learning paradigms that are considered in this paper are 1) that the expectation factor, and not just reinforcement, should be a major determinant of learning, and 2) that some stimuli can dominate and prevent learning to others by creating an expectation which neutralizes the reinforcement and prevents further learning.

Expectation as a Determinant of Effective Reinforcement

Since both reinforcement (activation) and expected reinforcement (expectation) determine their difference, and since it is this difference that controls the learning process, the theory predicts that the experimental correspondents of reinforcement and expectation will both effect the magnitude and direction of learning changes. In an instrumental conditioning paradigm the correspondent of modal reinforcement is the instrumental reinforcer. The correspondent of modal expectation is simply the presence or absence of stimuli that excite of inhibit the mode at the time the CR occurs, as these determine the activation level at this time. (The presence and vigor of the response might also be considered correspondents of the CR causing mode's activation level, but these are assumed to depend on relative rather than absolute activation levels) The learning process predicts that if the reinforcement exceeds the expectation, then those associations from modes highly active as the expectation was formed are increased, while if the reinforcement is less than the expectation, then these associations are decreased. Given the above correspondences between external events and reinforcement and expectation, this sort of result is well known to occur in instrumental conditioning experiments. The theory's explanations of the examples that follow will be much clearer if a few key assumptions are made explicit:

1. There is a normal nonzero base level of activation (reinforcement) for each of the mind's modes. Thus, changes to activation levels lower than this will be interpreted as negative effective reinforcement.

2. Responses can occur even when the corresponding response mode is at a low level of activation, as long as its competing modes are at as low or lower activation levels. Thus, a response may occur and be followed by an expectation of less reinforcement than the normal or base level if the whole mind is experiencing a below normal level of activation.

Global expectation level after a CR is determined by the instrumental reinforcers (including secondary reinforcers) present as the CR is performed. The global reinforcement level is determined by the instrumental reinforcers occurring after the CR is performed. Responses can occur with global expectations of relatively high or low reinforcement, and can be followed by relatively high or relatively low global reinforcement. This can be analyzed as a two dimensional continuum of possible combinations of global reinforcement after a response and of expectation as a result of the global reinforcement level as the response is performed. The experimental results of nine basic cases along this continuum are summarized in the box of figure 1.

GLOBAL REINFORCEMENT
(global activation level shortly after CR)

GLOBAL EXPECTATION (global reinforcement during CR)		HIGH	BASE	LOW
	HIGH	none during omission	– omission	– – combined effects
	BASE	+ appetitive conditioning	none normal affairs	– punishment conditioning
	LOW	+ + combined effects	+ escape, avoidance	none during avoidance

Figure 1. Direction of learning changes (effective reinforcement) and name or description of the experimental phenomena corresponding to nine basic cases of combinations of global expectation and global reinforcement.

Through the learning process the expectation of the CR mode will come to equal the global reinforcement resulting from the performance of the CR. The diagonal from the upper left to the lower right corners of the box corresponds to the cases where global expectation and global reinforcement following the CR are the same, before any learning, and thus none occurs. The case when both expectation and reinforcement are at base level corresponds simply to normal events - no reinforcement occurring, no expectations of reinforcement, and no instrumental learning. The other two cases on the diagonal, reinforcement and expectation high, and reinforcement and expectation low, also result in no learning. Some experimental paradigms that show this kind of a lack of learning, such as instrumental blocking and stimulus selection experiments, will be discussed in a later section.

All cells of the box below the diagonal correspond to cases of reinforcement exceeding expectation, and thus the development of a tendency for the CS to cause the CR (positive learning), while all cells below the diagonal correspond to cases of expectation exceeding reinforcement, resulting in the development of a tendency for the CS to prevent, or inhibit, the CR (negative learning). The CR causing mode becomes affected by the CS such that its expectation matches the reinforcement that is to follow. The middle horizontal row of cells corresponds to the simplest cases of instrumental conditioning, those with no unusual expectations involved. A reward (a positive instrumental reinforcer) following a response causes positive learning and a punishment (a negative instrumental reinforcer) causes negative learning.

The middle vertical row contains cases that have proved troublesome for some theories, as they strongly indicate the necessity for a concept of expectation. In these cases the effective reinforcement is simply the lack of an instrumental reinforcer. When the expectation is for punishment, then the absence of any instrumental reinforcer acts as a reward, as in escape or avoidance learning, while if the expectation is for reward, then its absence acts as punishment, as in omission training. The experimental manipulation that creates such expectations is the presentation of CS that has been paired with global reinforcement, i. e. that has been made a secondary reinforcer. This is precisely what has been hypothesized to affect global expectation of reinforcement. Thus, the single learning process does explain, or predict, the results as shown in the middle vertical row of the box.

In a typical escape conditioning experiment the subject is exposed to a CS (such as an auditory tone) that predicts the arrival of a negative instrumental reinforcer (a punishment, such as an electric shock), and thus acts as both a negative secondary reinforcer and an inducer of low expectation according to this paper's theory. The performance of the CR allows the animal to escape from the CS, providing a positive effective reinforcement. However, before the crucial CR is performed, but while the subject is subjected to the CS, other responses that the subject emits that are unrelated to the CS do <u>not</u> have their probability of occurrence strongly affected by the experience. The theoretical interpretation of this lack of learning is that while the CS causes a low reinforcement level, it has also been causing a low reinforcement level in the recent past, and thus has produced a low current expectation level. With reinforcement and expectation both low, the effective reinforcement, their difference, is very small or zero and there is no learning. This is the training situation referred to in the lower right cell of the box labelled "during avoidance." The situation with both reinforcement and expectation high, in the upper left cell, is directly analogous. In omission training the effective negative reinforcement occurs when the expectation of reinforcement caused by the CS ends. While the CS, creating both high reinforcement and high expectation, continues, there is no learning.

The other two cells that we have not individually considered, low expectation and high reward, and high expectation and low reward, are combinations of reinforcement and expectation influences that complement and enhance each other. These paradigms have not been studied as much (and have not been given names) because they seem to confound rather than separate the factors determining effective reinforcement.

The learning process that has been offered appears to be able to explain stimulus context expectation effects on the effectiveness of reinforcement in instrumental conditioning. The really interesting point, however, is that expectation phenomena in classical conditioning can be shown to result from the same learning process, only as it is used to explain classical conditioning.

In classical conditioning the correspondent of modal reinforcement is the arrival of the UCS. The correspondent of modal expectation is again the mode's level of activation as the CR occurs - or the sum of the effects of the active stimulus reflecting modes and their associations to the CR causing mode. As before, if reinforcement exceeds expectation the associations from the stimulus modes to the CR mode are increased while if expectation exceeds reinforcement they are decreased. An influence of expectation in determining effective reinforcement is clearly predicted for classical conditioning in addition to that which has been shown for instrumental conditioning. In an important paper Rescorla & Wagner convincingly argued from numerous experiments that reinforcement and nonreinforcement have varying effects on classically conditioned associations to CS's depending on the context of other stimuli within which the CS occurs (Rescorla & Wagner, 1972). Their theory has been well accepted and clearly captures the results of a large number of experiments (see Hilgard & Bower, 1975, p. 572). Rather than look at these individual experiments, the great similarity between Rescorla & Wagners' theory and the application of the theory presented here to classical conditioning will be demonstrated. It will be shown that the present theory makes the same qualitative predictions as the Rescorla & Wagner theory with one interesting exception, in which case the single learning process theory presented here fits the experimental data better than Rescorla & Wagners' theory.

Comparison of the Single Learning Process Theory for Classical Conditioning and the Rescorla/Wagner Theory

Rescorla & Wagners' 1972 paper shows that reinforcement and nonreinforcement have varying effects on classically conditioned associations formed to a CS depending on the context of other stimuli in which the CS occurs. For instance, if there is an association already formed between conditioned stimulus A and fear (reinforced by shock), then if stimulus X occurs at the same time as A does, and the combination is followed by shock, then there will be very little fear conditioned to X. Furthermore, if A has been conditioned to fear and then A and X occur together without being followed by shock, then X will actually come to be an inhibitor of fear. Numerous similar experiments led Rescorla & Wagner to construct a descriptive theory in which the animal takes account of what reinforcement is indicated by other stimuli when it adjusts the associative strength of a particular stimuli upon reinforcement or nonreinforcement. Stated cognitively, Rescorla & Wagners' theory is in perfect agreement with the theory presented here:

> "...organisms only learn when events violate their expectations. Certain expectations are built up about the events following a stimulus complex; expectations initiated by that complex and its component stimuli are then only modified when consequent events disagree with the composite expectation." (Rescorla & Wagner, 1972, p. 75)

From this idea and their numerous experiments Rescorla & Wagner then developed a more precise, more mathematical, statement of their theory. They formulated a theory in which the change in associative strength to each stimulus is a function of the difference between the reinforcement received on a trial and the total associative strength of the entire stimulus configuration. The essence of their theory as it will be considered here can be summarized in this equation giving the change in associative strength between a CS and the CR resulting from a classical conditioning trial (taken from Bolles, 1975, p. 164):

$$3) \quad \Delta Va = C \, (L - Vax)$$

where: C is a constant depending on the specific CS and US
L is the maximum associative strength with this US
ΔVa is the change in associative strength to stimulus A
Vax is the associative strength already present to stimulus A together with all other stimulus components X that may be present

Compare this with the equation for the single learning process with which this paper is concerned:

$$4) \quad \frac{d}{dt} \, {}_iV_j = {}_iC_j \, (A_j - P_j) \, {}_iE_j$$

where: ${}_iC_j$ is a constant depending on the specific modes i and j
A_j is the level of activation of mode j
${}_iV_j$ is the strength of the association from mode i to mode j
P_j is the expectation of activation of mode j
${}_iE_j$ is the eligibility of the association from mode i to mode j

Despite the differences in symbols, terminology, and listed meanings, there is an apparent gross similarity of form of the two learning equations. It will now be argued that this is a real and meaningful similarity, and its nature and extent will be investigated. The terms of the equations will now be taken one at a time and similarities and differences will be discussed.

iV_j, the strength of the association between the CS reflecting mode and the CR causing mode, is directly analogous to Va, Rescorla & Wagners' associative strength to a stimulus. For both of these measures of association between stimulus and response the presence of the stimulus increases the likelihood that the CR will occur if the association is positive and decreases this likelihood if the association is negative.

Rescorla & Wagners' total associative strength to a stimulus complex Vax is analogous to the effect on the CR producing mode of all associations from active modes, or the sum effect on this mode of present stimuli's associative strengths. This effect will be reflected in the activation level of the CR causing mode at the time of the CR, and in the expectation of this mode at the time of the UCS, or reinforcement, <u>as long as the overall effect is positive</u>. While the associative strength to a stimulus complex Vax is positive, it will be well represented as the expectation of the CR causing mode. If, however, this associative strength is negative, then the effect on the mode will be to decrease its activation, but its activation and thus its expectation later at the time of reinforcement can only approach zero - they can not go past it, they cannot be negative. Thus, the CR mode's expectation will not be a good analog for Vax, the total associative strength to a stimulus complex, when this total associative strength is negative. However, under these conditions expectation's value is not important because there will be very little learning. A negative associative strength to a stimulus complex means the complex decreases the likelihood that the CR will occur. If the total effect of the stimulus complex is to inhibit the CR producing mode, then it will reduce its level of activation, resulting in not only a low activation level and a low later expectation, but also a low later eligibility for learning changes. Thus, in those cases where expectation is important in determining learning changes, it <u>is</u> a reasonable analog for Rescorla & Wagners' Vax, the total associative strength to the presented stimulus complex. The multiplicative eligibility term iE_j of the single learning process formula has no analog in the Rescorla/Wagner theory. We will return to it and the differences between the two theories that it causes shortly.

The L term in the Rescorla/Wagner formula is said to be the asymptotic maximum level of associative strength, dependent on the particular reinforcer. This is directly analogous to the reinforcement A_j received by the CR producing mode, as this is the result of the intensity of the UCS's excitatory effect on this mode. The reinforcement determines how large the expectation can get, and thus the maximum associative strength it can support. Finally, the constant terms C and iC_j in the two formula clearly have identical roles and dependencies. Translating the single learning process into a difference equation for the result of a trial, and using analogous symbols, the following results:

5) $\Delta Va = C\ (\ L - Vax\ )\ iE_j^{*}$ if $Vax > 0$

Note that the eligibility term iE_j has been starred to indicate that it is not being used in its previous sense. To deal with the eligibility term iE_j in a way suited to the trial result form of the rest of the equation some additional reasoning is necessary. On all trials the CS occurs and the CS reflecting mode reaches a high level of activation. It is the product, or logical AND, of the activation levels of the CS reflecting and the CR causing modes at this time that determines the eligibility of the association later as the reinforcement arrives. Since the CS reflecting mode is always highly active, it follows that the eligibility will be determined primarily by the level of activation of the CR causing mode as the CR occurs. Thus this activation level will be a reasonable measure of eligibility to within a constant of proportionality (which can be included in the overall learning constant C). As before, this activation level is taken as a reasonable analog of of total associative strength to the stimulus complex Vax, when this strength is positive. Also as we determined before, when this strength is negative, the eligibility will be quite small and the learning process will change associations only a very little. To formalize and clarify this idea a little further consider a function f(X) to represent the activation level of a mode given its total input X, from associations from other modes and from external input. When X is significantly positive, f(X) will be approximately proportional to it, but as X becomes negative, f(X) remains positive and merely approaches zero (see figure 2).

Now the conversion of the single learning process's predictions to Rescorla & Wagners' form can be completed:

6) $\Delta Va = C \; (\; L - f(Vax) \;) \; f(Vax)$

This form has two differences from Rescorla & Wagners' equation (compare with equation 3). The most important one is the multiplicative f(Vax) term. The second difference, that of subtracting f(Vax) from L rather than just Vax is much less important because Vax and f(Vax) are very similar except when Vax gets near and less than zero, which is precisely the time when the f(Vax) multiplicative term goes to zero and makes the L-f(Vax) term unimportant, as the

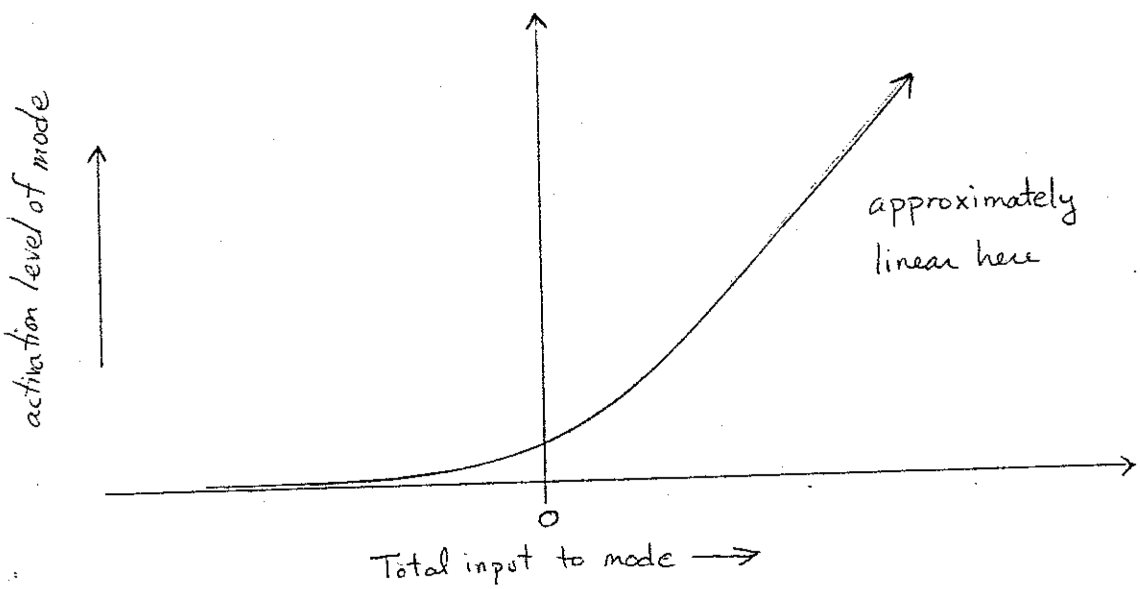

Figure 2. F(X), the relationship activation level and total input.

whole equation goes to zero.

Because of the multiplicative f(Vax) term, the single learning process's equation predicts that learning would be slow at first and then quickly pick up speed (initial positive acceleration). As the association between the CS and the CR becomes significantly large, the CS mode comes to excite the CR mode. This results in a higher eligibility of the association between them when the UCS (reinforcement) arrives, which in turn results in faster (accelerated) learning. Rescorla & Wagner's equation predicts only a negative acceleration in learning. As others have pointed out (Mackintosh, 1974, p. ii), their theory is in error here, as an initially positively accelerated classical conditioning learning curve is usually found (see figure 3).

This section has provided the first really powerful evidence for this single process explanation of classical and instrumental conditioning. After illustrating the way the theory handles the two basic conditioning paradigms, a certain characteristic of the learning process was taken up, namely the way learning is dependent on both a reinforcement and an expectation term. The consequences of this characteristic were computed for both conditioning paradigms, and both sets of phenomena were found in the experimental literature. The single learning process provides a unified theoretical way of viewing expectation as a determinant of reinforcement effectiveness in both

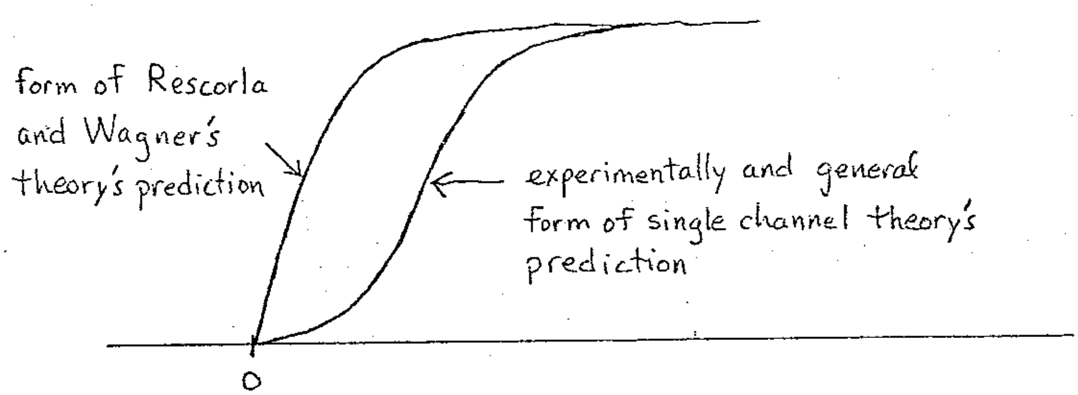

Figure 3. Associative strength in classical conditioning versus number of trials, as experimentally observed and as predicted by Rescorla and Wagner's theory and by the single channel theory.

classical and instrumental conditioning. The next section gathers more evidence for the single learning process theory by showing how it can provide theoretical explanation for another set of phenomena occurring in both kinds of conditioning.

Expectation as Stimulus Dominance Effects

Consider the generalized view of both kinds of conditioning that has been presented and developed in the previous sections. Each component stimulus of the stimulus complex contributes (positively or negatively) to the tendency to make the response and simultaneously to the upcoming expectation of reinforcement. If these conditioned stimuli (CSs) produce too great a combined response tendency then they create a greater expectation than the reinforcement that will be received and the learning process reduces the associations from the CSs. If the reinforcement is of a fixed size, then there will be a fixed maximum total response tendency that can be learned to the CSs and which will be stable. This suggests that the single learning process might provide some explanation for the many phenomena in both classical and instrumental conditioning in which learning to one CS dominates other CSs and prevents or inhibits learning to them. Typical instances are blocking, in which prior learning to one CS prevents any learning to a second CS presented simultaneously with the first, and overshadowing, which is a general tendency for all response tendency to be allotted to one particular aspect of the distinguishing stimulus configuration, even though all aspects are equally informative. This section will show that the single learning process theory, with no further assumptions, can indeed explain such phenomena as they occur in both classical and instrumental conditioning. This single process explanation of a major class of phenomena that occurs in both kinds of conditioning strengthens the evidence that this theory captures important aspects of the relationship between the conditioning paradigms.

The set of stimulus dominance phenomena that occur in both conditioning paradigms can be divided into three classes based on what aspects of the stimulus causes it to dominate the others. The three stimulus aspects are greater salience, prior learning, and greater validity. This section will concern itself primarily with the explanations of the stimulus dominance phenomena involving <u>instrumental</u> conditioning. The established work of Rescorla & Wagner can be relied on for stimulus dominance phenomena in classical conditioning. It has been well documented (Rescorla & Wagner, 1972; Hilgard & Bower, 1975, p. 572) how their theory, which has been shown to give essentially the same predictions as the present theory's for classical conditioning, can explain each of the phenomena that will be discussed.

The experimental situation in instrumental conditioning that we are interested in is instrumental discrimination learning. The subject is reinforced for emitting the response in the presence of some CSs, the S+s, but not for emitting it in the presence of other CSs, the S-s. This situation is very similar to classical discrimination learning. The UCS in classical discrimination learning is analogous to the reinforcement in the instrumental case, and follows only S+ CSs, not S- CSs. The same sort of analysis applies to both kinds of discrimination learning, and it is essentially a generalized version of that which has been given by Rescorla & Wagner. The S+ produce an expectation of the reinforcement (of the UCS in classical conditioning) and learning proceeds according to the difference between the expectation and the reinforcement. When the S+ is a collection of stimuli ABC... there are a number of reasons why the association to the response may be to only one of them. Recall that all associations contribute to the response causing modes expectation. Since this expectation cannot exceed the reinforcement, the total association to all the S+ is limited. Suppose the constant iC_j was very large for the association from A, an S+, to the response relative to the constants for the associations from the other S+s BC... to the response. The association to the response from A, with its high iC_j, would change much faster than the associations from the other stimuli if they were presented identically. Thus, when total expectation reaches the reinforcement level, the primary contributor will be A, the S+ with the large constant iC_j. Experimentally, in both classical and instrumental conditioning, this phenomena, known as overshadowing, frequently occurs (Sutherland & Mackintosh, 1971, pp. 144-153). Even though many stimulus aspects indicate that the UCS is about to occur, or that the reinforcement will come if the CR is made now, typically only one has had the majority of the CR making tendency associated with it (and the others are less associated with the CR than if this dominating CS was not present). Which stimulus aspect becomes dominant can be affected by varying the relative intensity of the stimulus aspects. Increasing its intensity will make a stimulus aspect more likely to become dominant. According to the above developed classification this would be called stimulus dominance by salience.

A stimulus can also become dominant because of being involved in prior learning. If stimulus A, an S+, has been fully trained and associated with the CR, then it creates an amount of expectation equal to the reinforcement used. If another stimulus B is made to accompany A, and thus be an equally informative S+, then it will have no learning changes made to its association to the CR because A already fully predicts the reinforcement, and there can be learning only when reinforcement differs from expectation. This phenomena, known as blocking, is known to occur in both classical and instrumental conditioning (Sutherland & Mackintosh, 1971, pp. 110-118). Training with stimulus A as an S+ will prevent any further learning on training trials using A and B together as an S+.

Finally, stimuli can become dominant by virtue of having greater discriminative validity. If stimulus A always indicates that reinforcement will follow the response (or that the UCS is about to occur), while stimulus B is an imperfect indicator (i. e. is not always correct), then eventually, whatever the constants iC_j, and whatever the initial response tendencies to A and B, there will be no response tendency to B - it will be completely dominated by A. This is because once the combination of response tendencies to A and B has reached the level of the reinforcement, then each time A occurs without B the response tendency to A will increase some. When A and B occur together again, the combined expectation is too high, and both A and B are decreased in their response tendencies, with a net gain for stimulus A. On the other hand, each time B occurs without A, the reinforcement fails to occur, so if there is any expectation generated by B, it will be decreased on such a trial. When A and B occur together again the combined expectation will be too low, and both A and B are increased in their associations to the CR, again with a net gain for stimulus A. These processes will continue until the response tendency to A produces exactly all the expectation to match the reinforcement, and the response tendency to B is zero. A nice series of experiments by Wagner et. al. has demonstrated precisely these results for both classical and instrumental conditioning (Wagner et. al., 1968).

SUMMARY AND CONCLUSIONS

A very important question throughout the history of learning theory has been the nature of the relationship between the two basic training procedures, that of classical and that of instrumental conditioning. Many theories have tried to unify the two forms of conditioning by explaining one in terms of the other with little success. The present theory claims to explain classical and instrumental conditioning not as forms of one another, but as different aspects of a single learning process which can not be identified exactly with either.

A model of the mind as consisting of modes in various states of activation reflecting the presence of stimuli and causing overt responses was developed. The single learning process was stated as affecting associations between modes that determine how they influence each others level of activation. It was illustrated how this theory could explain the basic results of the experimental procedures of classical conditioning, instrumental conditioning, and of the creation of secondary reinforcers. A neuronal basis for the theory was described, linking the present theory with previous neuron-oriented work.

By making the identification of activation of a mode and reinforcement of the mode, and of recent activation level of a mode and expected, or predicted, activation (reinforcement) level of the mode, some interesting interpretations of the learning process were made. Since it was claimed that it was the difference between reinforcement and expected level of reinforcement that determined the effective reinforcement, it was noted that there should be ways to alter the effectiveness of a reinforcer by altering the context of expected reinforcement, and that this should occur in both classical and instrumental conditioning. Such ways were found to be known to occur in both conditioning paradigms, and they were shown to be consistent with the single learning process and its explanations of the conditioning paradigms. This was done in classical conditioning by showing the similarity of the present theory to a descriptive theory of classical conditioning by Rescorla & Wagner. The theories produce the same qualitative predictions except for the form of the acquisition curve, where the Rescorla/Wagner theory does not fit the experimental results as well as the single process theory presented in this paper. The effect of context on the effective magnitude of reinforcement in instrumental conditioning was established by a brief review and classification of the many instrumental conditioning phenomena demonstrating such an effect.

The single learning process presented in this paper implies that as long as the reinforcement received by a mode was predicted by the stimuli that preceded it there should be no learning. Thus, in certain situations it should happen that some stimuli completely predict the reinforcement and dominate other stimuli in that they prevent any association between the others and the CR. The conditions under which this can be expected to occur, given the hypothesized single learning process, were listed and identified with important experimentally observed phenomena, each of which occurs in both classical and instrumental conditioning.

The theory presented here has been able to interpret two general important sets of phenomena, context effects on effective reinforcement and stimulus dominance effects, present in both classical and instrumental conditioning, as different aspects of a single learning process, and agree with experiment as well or better than current separate theories of these sets of phenomena. This constitutes strong evidence that the theory captures important aspects of the relationship between classical and instrumental conditioning.

Only a few learning phenomena have been considered in this paper to demonstrate the theory's unified treatment of the two conditioning paradigms. An interesting and important question is whether this single process theory can be extended to other learning phenomena. The answer is probably in some cases yes, and in some cases no. This kind of single process theory seems to the author to be adequate with respect to most other learning phenomena and especially useful for some, notably instrumental latent learning and sensory preconditioning (Sutton, unpublished). However, the theory appears to need significant modification or extension to satisfactorily explain some important learning phenomena, such as contrast effects. Thus, although the evidence presented in this paper does strongly indicate that this theory captures important aspects of the relationship between classical and instrumental conditioning, it is admittedly not a finished product.

REFERENCES

Bolles, R. C. Learning Theory. New York: Holt, Rinehart and Winston, 1975.

Griffith, V. V. A Mathematical Model of the Plastic Neuron. Goodyear Aircraft Corporation, Akron, Ohio, Rept. No. GER-10589, 1962.

Griffith, V. V. A model of the plastic neuron, IEEE Trans. on Military Electronics MIL-7, 1963, 243-253.

Hilgard, E. R. & Bower, G. H. Theories of Learning, fourth edition. Englewood Cliffs, N. J.: Prentice-Hall, 1975.

Klopf, A. H. Brain Function and Adaptive Systems - A Heterostatic Theory, Air Force Cambridge Research Laboratory Special Report No. 133 (AFCRL-72-0164), 1972, L. G. Hanscom Field, Bedford, Mass.

Mackintosh, N. J. The Psychology of Animal Learning. New York: Academic Press, 1974.

Olson, G. M. An information processing analysis of visual memory and habituation in infants. In Tighe, T. J., & Leaton, R. N. (Eds.), Habituation: Perspectives from child development, animal behavior, and neurophysiology.

Hillside, N.J.: Lawrence Earlbaum Associates, 1976.

Rescorla, R. A. & Solomon, R. L. Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning, Psychological Review, 1967, 74, 151-182.

Rescorla, R. A. & Wagner, A. R. A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black and W. F. Prokasy (Eds.), Classical Conditioning: II Current Research and Theory. New York: Appleton, 1972.

Sutherland, N. S., & Mackintosh, N. J. Mechanisms of Animal Discrimination Learning. New York: Academic Press, 1971.

Sutton, R. S. Learning theory support for a single channel theory of the brain, unpublished.

Wagner, R. A., Logan, F. A., Haberlandt, K., & Price, T. Stimulus selection in animal discrimination learning, Journal of Experimental Psychology, 1968, 76, 171-180.

Wilkins, M. G. Neural Modelling: Methodology, Techniques and a Multilinear Model for Information Processing, Biological Computer Laboratory Technical Report No. 19, 1970, University of Illinois, Urbana, Illinois (DDC Report

AD 727770).