A Model for Stimulus Generalization in Pavlovian Conditioning

John M. Pearce
University College, Cardiff, Great Britain

A selective review of experiments that can be said to demonstrate the effects of generalization decrement in Pavlovian condition is presented, and it is argued that an adequate theoretical explanation for them is currently not available. This article then develops a theoretical account for the processes of generalization and generalization decrement in Pavlovian conditioning. It assumes that animals represent their environment by a stimulus array in a buffer and that this array in its entirety constitutes the conditioned stimulus. Generalization is then held to occur whenever at least some of the stimuli represented in the array on a test trial are the same as at least some of those represented in the array during training. Specifically, the magnitude of generalization is determined by the proportion of the array occupied by these common stimuli during training compared to the proportion of the array they occupy during testing. By adding to this principle rules concerning excitatory and inhibitory learning, it is proposed, the model can explain all the results that were difficult for its predecessors to account for.

A fundamental property of conditioned responding is that it may occur in circumstances that differ from those prevailing during acquisition. This phenomenon is referred to as generalization, and the term generalization decrement is used when this transfer is less than complete. These effects have been revealed with both Pavlovian and instrumental conditioning, using a variety of species and a wide range of conditioned stimuli (CS) and unconditioned stimuli (US).

As far as instrumental conditioning is concerned, a considerable number of studies have been directed at examining the factors that influence the magnitude of generalization from one situation to another (for reviews see Honig & Urcuioli, 1981; Mackintosh, 1974). These results also have been the focus of considerable theoretical analysis. In contrast, much less attention has been paid to the role of generalization in Pavlovian conditioning. Only a few studies have been concerned specifically with demonstrating the generalization decrement of a Pavlovian conditioned response (CR) as a result of modifying the properties of the CS (Hoffman & Fleshler, 1961; Holvand, 1937; Moore, 1972; Pavlov, 1927; Siegel, Hearst, George, & O'Neal, 1968). Furthermore, only a few theorists have been specifically concerned with the analysis of these and related effects (Hull, 1943; Pavlov, 1927). An alternative approach has been to show how a particular theoretical analysis of the Pavlovian learning process can be extended to account for generalization and generalization decrement (cf. Rescorla, 1976). The argument presented in this article is based on the premise that an adequate account for the effects of generalization must lie at the core of any theory of Pavlovian learning. For example, most Pavlovian conditioning studies involve a test stage that is conducted in a different manner to that employed during training. Unless it is known to what extent the test results are due to the effects of generalization decrement consequent upon this change in procedure, it will be difficult to assess the influence of any other process that may be effective in the experiment.

In this first part of this article, a variety of demonstrations of stimulus generalization and generalization decrement are described. The extent to which these results can be explained by a number of theories of learning is also examined. It is argued that although these theories can provide a reasonable explanation for most of the results, none can provide a satisfactory explanation for all of them. In the second and major part of this article, a novel theoretical interpretation of these effects is presented and is used to explain a variety of additional results from conditioning experiments. Most of the discussion is concerned with Pavlovian conditioning. This is not to imply that the effects described or the interpretation offered do not apply to instrumental conditioning. Indeed, it would be surprising if much of what is said here did not apply to this procedure. It is principally in order to confine this article to a reasonable length that it is restricted to Pavlovian conditioning.

Evaluation of Theories of Stimulus Generalization

Two basic procedures have been used to demonstrate the effects of stimulus generalization in Pavlovian conditioning. The more straightforward of these is to train subjects with a single CS, such as a tone, and then during a test phase to examine the strength of the CR elicited by stimuli that differ from the CS in some way—for example, in frequency. In a series of studies based on this design, Moore (1972) has shown that the strength of the CR elicited by a test CS is inversely related to the magnitude of the physical difference between the training and test stimuli. This finding replicates many similar results.

I should like to thank Helen Kaye, Douglas Young, Anthony Dickenson, and R. A. Rescorla for their helpful suggestions concerning the ideas presented in this article.

Correspondence concerning this article should be addressed to John M. Pearce, Department of Psychology, University College, Cardiff CF1 1XL, Great Britain.
from operant experiments and has been demonstrated on a number of occasions (Hovland, 1937; Pavlov, 1927; Siegel et al., 1968).

The second procedure for examining stimulus generalization differs from the first in that the physical properties of the trained CS remain unaltered for testing. Instead, the stimuli accompanying the CS are changed in some way. It is demonstrations of generalization, and in particular generalization decrement, with this procedure that provide the greater difficulty for theoretical analysis.

Overshadowing (Pavlov, 1927) and blocking (Kamin, 1969) are two procedures that demonstrate effects that may be described as generalization decrement with the latter method. In both, subjects receive conditioning with a compound CS, AB; prior to being presented with one of the elements, B, alone. Overshadowing is said to be successful if this training results in a weaker CR being elicited by B than if subjects were conditioned with that stimulus in isolation. In blocking experiments, subjects receive conditioning with A prior to compound training. This is found to result in B eliciting a weaker CR than when pretraining with A is omitted. Thus, the generalization decrement induced by the removal of A from the AB compound can be enhanced by pretraining with A. This finding is particularly important because it is inconsistent with a number of the early attempts to account for stimulus generalization in Pavlovian conditioning.

Hull's Theory

Hull (1943, 1952), for example, maintained that the occurrence of a CS by itself resulted in the formation of an internal representation, S, that can acquire response eliciting properties as a result of Pavlovian conditioning. If this stimulus is presented in compound with another CS, its internal representation will be transformed to S' according to the principles of afferent neural interaction. Thus in overshadowing, subjects will be trained with a target CS with an internal representation, S', and tested with the same stimulus evoking the internal representation S. These representations will be different, and the greater this difference the weaker will be the CR on testing. Provided that blocking and overshadowing use the same stimuli, the transition from compound conditioning to testing the target CS by itself will be the same in both cases. As a result, the transformation of the representation of the CS from S' to S will also be the same and the decrement observed in the strength of the CR on testing should be identical for both procedures. The frequent finding that blocking results in a greater decrement of responding than overshadowing is therefore inconsistent with Hull's analysis and suggests that some alternative, or additional, mechanism to afferent neural interaction is responsible for blocking. Unfortunately, it is not at all clear from Hull's theorizing what this mechanism might be.

Rescorla–Wagner Model

In contrast to the approach of Hull, the Rescorla–Wagner (1972) model (see also Wagner & Rescorla, 1972) can readily explain the different outcomes of overshadowing and blocking. Indeed, it was developed principally to account for these and related findings. According to this model, the properties of the unconditioned stimulus (US) determine the asymptote of conditioning, and stimuli will continue to gain in associative strength until their aggregate strength is equal to this asymptote. This proposal is expressed by Equation 1 (Rescorla & Wagner, 1972; Wagner & Rescorla, 1972).

\[ \Delta V_A = \alpha \beta (\lambda - V_I) \]  

(1)

\[ \Delta V_A \] is the associative strength acquired by A on any conditioning trial that this stimulus is present. \[ V_I \] is the total associative strength of all the stimuli present on that trial, and \( \lambda \) is the asymptote of conditioning determined by the US. The learning rate parameters \( \alpha \) and \( \beta \) have values between zero and one and are related to properties of, respectively, the CS and US. In blocking in which one stimulus, A, is used to signal a US during pretraining, the associative strength of A, \( V_A \), will eventually equal \( \lambda \). When a second stimulus, B, is then presented with A for compound conditioning, the value of \( V_I \) will be given by the sum \( V_A + V_B \). From the first conditioning trial this will be equal to \( \lambda \), and according to Equation 1 there will be no increase in the value of \( V_B \). Thus, according to this analysis, the gain in associative strength to the added element during blocking should be negligible. A different outcome is predicted when this model is applied to overshadowing. At the outset of compound conditioning, both stimuli will be novel, and \( V_I \) can be assumed to equal zero. With training, \( V_A \) and \( V_B \) will gain in associative strength until their combined strength is equal to \( \lambda \). At this point there will be no further changes in associative strength, and the elements will separately elicit a weaker CR than if they had been reinforced in isolation. The strength of this CR should nonetheless be greater than that acquired by the added element of a blocking study.

Blough (1975) has shown how this model can be extended to account for stimulus generalization when the properties of a CS are transformed. Specifically, he proposed that individual stimuli such as lights and tones are represented by elements that may be activated to varying degrees. Each element is assumed to be capable of acquiring associative strength and the influence of this on performance is held to be directly related to its degree of activation. Changing the properties of a stimulus, such as its frequency in the case of an auditory CS, will result in a different set of elements being activated. Some of these will be the same as the elements activated by the original CS but to a lesser extent and, as a consequence, this procedural change will result in a decrement in the strength of the CR.

The account provided by Blough (1975) is extremely successful in predicting the outcome of generalization studies when the CS itself undergoes some change. When it is applied to the results considered below, however, in which generalization decrement is revealed by manipulating the stimuli accompanying the CS, it encounters the same difficulties that will be shown to confront the Rescorla–Wagner model. This stems from Blough’s (1975, p. 19) adherence to Equation 1 as a description of the course of conditioning. Thus, the term \( V_I \) is still used to represent the combined associative strength of all the stimuli present on a trial, but in Blough’s account the associative strength of each stimulus is determined by the sum of the associative strengths of all the elements it activates.

There are at least three different findings that provide results
inconsistent with the Rescorla–Wagner model as it has been presented. These are also inconsistent with the version proposed by Blough (1975). The first concerns the prediction that overshadowing should not be evident after only one compound conditioning trial. According to Equation 1, the discrepancy \((\lambda - V_T)\) determines the increase in associative strength on a particular trial to a given CS. On the first overshadowing trial the value of this discrepancy will be the same as for subjects trained with a single CS. Consequently, the gain in associative strength should be the same whether a stimulus is reinforced in compound or in isolation. There is, however, ample evidence demonstrating that one compound conditioning trial is sufficient to produce overshadowing (James & Wagner, 1980; Mackintosh, 1971; Mackintosh & Reese, 1979). Mackintosh (1971), to cite one example, demonstrated that if a single conditioning trial was conducted with a light–tone compound, the strength of the CR elicited by the tone was considerably less than for subjects receiving one conditioning trial with the tone alone.

The second procedure to produce results that are difficult to explain in terms of the Rescorla–Wagner model has been referred to as the feature-positive design. In this, subjects receive nonreinforced presentations of one stimulus, \(A^o\), intermixed among reinforced presentations of a compound, \(AB^+\). With sufficient training, the occurrence of \(A\) will not elicit a CR, whereas \(AB\) will elicit a strong CR. In a series of experiments involving this procedure, Young and Pearce (1984) examined the effects of presenting the feature CS, B, in the absence of the nonreinforced element. When B was presented by itself after discrimination training, it elicited a considerably weaker CR than that performed in the presence of the compound. In other words, the presentation of B in the absence of A resulted in an effect that can be described as generalization decrement. This outcome is not anticipated by the Rescorla–Wagner model, because it follows from Equation 1 that such feature-positive training will reach asymptote when \(V_A = 0\) and \(V_B = \lambda\). At this point, performance in the presence of B should be the same whether or not A is present. It might be argued that the failure of Young and Pearce (1984) to confirm this prediction was due to the incompleteness of discrimination training and that A retained a residue of associative strength that augmented the properties of B on compound trials. We could find no evidence to support this argument, however (Young & Pearce, 1984).

The third finding to be considered involves the phenomenon referred to by Pavlov (1927) as external inhibition. If subjects are trained with a target CS prior to its being presented in compound with an associatively neutral stimulus, this can result in a weakening of the CR elicited by the target stimulus. Thus, Kamin (1969, p. 49) reported on the first compound conditioning trial of a blocking experiment that the CR elicited by the pretrained element was somewhat weaker than on the previous trials when this stimulus was presented by itself (see also Kamin & Szakmary, 1977). Once again the implication of these results is that changing the conditions from training to testing, this time by adding a stimulus, will reduce the strength of previously acquired conditioned responding. The problem posed by this sort of result for the Rescorla–Wagner model is that because it assumes that the strength of a CR is determined by the sum of the associative strengths of all the stimuli that are present on a trial, it follows that the addition of a neutral stimulus to a CS should not influence the performance of a CR. The finding that there is a decrement in the strength of a CR when a CS is first presented in compound with a neutral stimulus is inconsistent with this aspect of the theory.

As it has been presented, the Rescorla–Wagner (1972) model is unable to explain the results of experiments in which subjects are required to respond differently to a compound and its constituents. Positive, \(A^oAB^+\), and negative, \(A^oAB^B\), patternings are two examples of such procedures that nonetheless can be successfully mastered (e.g., Bellingham, Gillette-Bellingham, & Kehoe, 1985; Woodbury, 1943). Wagner and Rescorla (1972, p. 306) were fully aware of the problem and proposed that the presentation of a compound results in the formation of a hypothetical “configural” stimulus that is absent when the elements are presented alone. Thus, with positive patterning, the occurrence of a CR in the presence of the compound would be attributed to the configural stimulus acquiring strength, whereas, with a negative pattern, the acquisition of inhibitory strength by this stimulus would be held responsible for the weaker CR in the presence of the compound than the elements.

It is conceivable, in the feature positive experiment (\(A^oAB^+\)) by Young and Pearce (1984), that the configural stimulus generated by the AB compound acquired a significant measure of associative strength. As a consequence, the presentation of B by itself would result in a weaker CR than that observed to the compound because of the absence of the hypothetical element. Thus, once it is acknowledged that compounds can generate hypothetical stimuli, some of the results cited above are no longer incompatible with the Rescorla–Wagner (1972) model. But even when appeal is made to a configural stimulus there still remain some results that do not fit within this framework. For instance there is no reason according to Equation 1 that the presence of a configural stimulus should be responsible for the effects observed in a one-trial overshadowing experiment. It is also impossible to account for external inhibition by referring to such a stimulus.

From the perspective of the present discussion, a cardinal feature of the Rescorla–Wagner model is its assumption that the associative strength of a CS is unaffected by the context in which it is presented—an assumption that is not unique to this model. During the last 10 years or so a number of theories of Pavlovian conditioning have been proposed that differ considerably but that all maintain that the CR elicited by a CS will be unaffected by the addition or removal of stimuli that can be assumed to be associatively neutral (Frey & Sears, 1978; Mackintosh, 1975; Moore & Stickney, 1980; Pearce & Hall, 1980). It is worth noting that the results that have proved a problem for the Rescorla–Wagner model are also difficult for these theories to explain. As noted by Pearce and Hall (1980, p. 541), for example, demonstrations of overshadowing after one compound trial cannot be explained by these theories. In addition, they are all contradicted by successful demonstrations of external inhibition. In view of these common difficulties it is tempting to speculate that they all stem from the same erroneous common assumption. That is, they are incorrect in assuming that the associative strength of a CS is unaffected by any change in the stimuli accompanying its presentation. The model that is described shortly does not make this underlying assumption.
Configural Theories

The theories considered thus far all assume that the pattern of stimulation on a trial can be broken down into elements, each of which is capable of entering into an association with the US. In contrast, a number of authors have proposed that such a pattern cannot be broken down in this way but instead itself constitutes a unique CS. For instance, in the case of positive patterning, a representation of the AB compound in its entirety might be associated with the US, whereas separate representations of A and B would not be associated with this event.

One of the earliest attempts to formalize such a configural view was by Gulliksen and Wolfe (1938), who were concerned with instrumental discrimination learning. Suppose that subjects have been trained to approach a black door and to avoid a white one, irrespective of their position in a jumping stand. According to Gulliksen and Wolfe (1938), this was not achieved by learning to approach a specific stimulus, but instead depended upon subjects learning to go left in the presence of a particular configuration (black door on the left, white on the right) and to go right in the presence of the converse configuration. As it is developed, the model is concerned principally with discrimination learning in which the discriminanda vary from one another along dimensions such as luminance. Thus, it is not clear how it should be applied to experiments in which an entire dimension is removed after training, as with overshadowing; or why overshadowing and blocking should lead to different outcomes.

The mixed model of conditioning described by Atkinson and Estes (1963, p. 243) can also be described as a configural model of learning (see also Estes & Hopkins, 1961; Friedman & Gelfand, 1964). According to this model, the stimuli present on a conditioning trial are represented by a set of elements of which a sample serves as the CS. On any trial the entire sample that is selected will be associated, as a unit, with the event that occurs in its presence. There is no reason that samples with common elements should not be associated with different events, so that for positive patterning a sample representing the AB compound will be associated with the US, whereas the samples selected in the presence of A or B alone will not.

To account for stimulus generalization it is held that when a novel stimulus is presented, it will elicit a response of magnitude and nature determined by the proportion of elements that it has in common with previously conditioned samples. After a conditioning trial, however, the events of that trial will determine exclusively the nature of the response that is elicited by the sample on subsequent trials. This last assumption is important to the present discussion because it leads the mixed model into predicting, incorrectly, that blocking and overshadowing should exert a similar influence. When training with A is conducted prior to conditioning with AB, then on the first compound trial, generalization from A to AB will result in a CR. But as a result of the conditioning with AB, the sample representing this compound will itself acquire the capacity to elicit a CR of asymptotic strength. If B is then presented alone, the strength of the CR generalizing to it from AB will be the same as if previous training had been conducted only with AB.

According to Medin (1975; see also Medin & Reynolds, 1985; Medin & Schaffer, 1978), all the stimuli on a trial, including those provided by the experimental context, activate a cue-context node that enters into an association with the event with which it is paired. On subsequent trials, the strength of a response is determined by the extent to which the node is activated, which in turn is determined by the similarity of the stimuli present during testing with those during training. To determine the similarity between two stimuli it is assumed that any stimulus is represented by a number of dimensions, and the similarity of one stimulus to another along a single dimension is given by a parameter that varies in value between 0 and 1. The larger the value of this parameter the greater the similarity on that dimension. The parameters for the various dimensions that the two stimuli have in common are then presumed to combine multiplicatively to determine their overall similarity, which in turn will determine the degree of generalization between them.

The theory has been developed to deal successfully with a range of instrumental discrimination tasks, but as Medin (1975, p. 302) points out, it has not been fully developed as far as Pavlovian conditioning is concerned, and in this respect it is difficult to derive precise predictions from the theory. Indeed, as it is presented at one point it appears as if compound conditioning should not result in any of the elements eliciting a response when they are presented in isolation. This is based on Medin's (1975, p. 270) proposals that (a) when two stimuli are presented together the presence of one will serve as the context for the other and (b) that both the stimulus and the context must be represented if the cue-context node is to be activated. Removing one of the elements from a compound CS will therefore remove one of the representations essential for the activation of the appropriate node, and a CR should not be recorded.

Another problem posed by results from Pavlovian conditioning for this theory concerns external inhibition. As it is formulated, it is not clear that the combination of a neutral stimulus with a CS should reduce the extent to which the latter activates its cue-context node. This would have to be the case if the theory is to account for the weakening of the CR that can be observed with this procedure.

Finally, an obstacle for all the configural theories considered in this section is that they have little to say about inhibitory conditioning. It is now well established that nonreinforced presentations of a compound comprising a neutral stimulus and an excitatory CS will result in the former acquiring negative or inhibitory associative strength (Pearce & Hall, 1980; Wagner & Rescorla, 1972). A stimulus with such associative strength can elicit its own CRs, reduce the capacity of any CS with which it is paired to elicit CRs, and enter only slowly into associations when paired with a US (see Lolordo & Fairless, 1985, for a recent discussion of this topic). For all the configural theories considered here, the effect of nonreinforcement is simply to weaken the effects of previous conditioning, and they are therefore unable to explain these effects. This failure to provide a satisfactory account of inhibitory learning, together with the other problems that have been identified, means the current configural theories provide an inadequate account of Pavlovian conditioning.

The Model

The basic feature of the model has much in common in Atkinson and Estes' (1963) mixed model of conditioning. It pre-
sumes that animals possess a buffer of limited capacity that is always full and represents the environment or overall pattern of stimulation to which the subject is currently exposed. For associative learning, the main principle of the model is that whenever a representation in the buffer is followed by a US, this representation in its entirety will serve as the CS. Thus, as a result of Pavlovian conditioning, subjects are assumed to form in long-term memory (LTM) a representation of the entire contents of the buffer that were present prior to US onset. The strength of this associative bond is assumed to grow with repeated conditioning trials and to influence directly the strength of the CR.

If contents of the buffer should change, even slightly, from one trial to the next, a new CS-US association will be formed in LTM. On any trial the contents of the buffer are compared with the CS representations in LTM. In some cases there will be a perfect match between the two; on other occasions they may be quite different. The greater the degree of this similarity the stronger will be the influence of the CS-US association in determining the strength of the CR that is elicited. A more formal account of these proposals is presented below, starting with a discussion of generalization, followed by a model of conditioning.

Generalization

Whenever a stimulus is presented, it will excite a particular pattern in the buffer that will also include representations of the events accompanying that stimulus. Changing the stimulus in some way—for example, if it is a tone increasing its frequency from T to T′—will result in a change in the buffer contents to a certain extent. As with Medin’s (1975) theory, a parameter, S, which can vary between 0 and 1, is used to represent similarity across different trials. If the two tones just referred to are very similar, the value of this parameter, S_T, will be close to 1, but if they are very different then its value will be closer to 0. An important point to note—and this is where the model differs from Medin’s account—is that even if the stimulus is changed entirely (for example, the tone is replaced by a light, L), the parameter S_L will still have a positive value. This follows from the assumption that the array on both trials will contain representations of stimuli that remain constant, such as those provided by the experimental context, and these will ensure a measure of similarity between the arrays when the light and the tone are presented.

The importance of the parameter S lies in the fact that it is used to determine the extent to which the effects of conditioning will generalize from one CS to another. Thus, if E_A represents the excitatory strength of CS A, Equation 2 indicates the extent to which this strength will generalize to a similar stimulus A′, E_A′. It will become evident that a critical and novel feature of the model is that it distinguishes between excitatory strength that is conditioned to a stimulus (E) and the one that generalizes to it (e).

\[ e_A = \alpha S_A \cdot E_A \]  

(2)

In view of the role played by S in the model, it is necessary to indicate how its value may be determined. It is assumed that the intensity of a stimulus determines the area it is afforded in the buffer, so that the greater the intensity of an event the more space it will occupy. Because the buffer is of fixed capacity, the absolute intensity of a stimulus will not determine directly the area it occupies. Instead, it will be determined by the ratio of the intensity of the stimulus in question to the total intensity of all the stimuli present on a trial. Equation 3 is based on these principles and indicates how the similarity of the buffer contents can be computed for trials when A and A′ are presented.

\[ \alpha S_A = \frac{P_{\text{com}}}{P_{ZA}} \times \frac{P_{\text{com}}}{P_{ZA'}} \]  

(3)

The value of \( P_{\text{com}} \) is determined by the perceived intensity (P) of the stimuli that are in common on both occasions. In the present example, this will include contextual stimuli as well as any elements that are common to A and A′. The values of \( P_{ZA} \) and \( P_{ZA'} \) are set according to the total perceived intensity of stimulation on trials when A and A′, respectively, are presented. Thus, the more similar are the buffer contents on the two trials, the closer will \( P_{\text{com}} \) be to \( P_{ZA} \) and \( P_{ZA'} \), and the higher will be the value of \( \alpha S_A \).

For S to vary when a stimulus is changed in some respect, it must be assumed, in a manner similar to that suggested by Blough (1975), that its representation in the buffer consists of a distribution of activated elements. The intensity of the stimulus will determine the number of elements that are activated. Changing the stimulus will alter the distribution of elements in such a way that the greater the change the fewer will be the common elements that are activated. As a result, the value of S will be relatively high if the stimulus change is slight, but low if this change is considerable. Equations 2 and 3, therefore, correctly lead to the prediction that the strength of the CR generalizing from one CS to another is a function of the difference between them.

Very little more will be said about generalization that occurs when the properties of a stimulus are changed. Instead, the discussion now focuses on the way the above principles can be applied to procedures in which the contents of the buffer are changed either by the removal of a stimulus, as in the case of overshadowing, or by addition of a stimulus, as in the case of external inhibition and overexpectation.

Application to overshadowing: It is a relatively simple matter to develop the above principles to explain why conditioning with a compound CS may often result in a relatively weak CR being elicited by at least one of its elements when subsequently presented in isolation. Because the S parameter is determined by the similarity of the stimuli present during training and testing, it follows that its value will determine the magnitude of generalization across these trials. Thus, if subjects receive excitatory conditioning with a tone-light compound, and are then

---

1 As Equation 2 stands, there is a direct relationship between the similarity of two stimuli and the magnitude of generalization between them, but this may not be correct. It is conceivable that an S-shaped function exists between the amount of generalization and similarity, so that with high values of S, Equation 2 underpredicts the degree of generalization, whereas the converse is true for low values of S. One reason for making such a proposal is that it is consistent with the general view that the shape of the generalization gradient is bell shaped, rather than linear.
tested with either the light or the tone, the excitatory strength
generalizing to the tone will be given by Equation 4a and that
to the light by Equation 4b:

\[ e_T = \frac{TL_{ST} \cdot E_T}{P_{TC}} \]  
\[ e_L = \frac{TL_{SL} \cdot E_T}{P_{TC}} \] (a) (4)

(E_T is the excitatory associative strength of the light–tone com-
 pound, \( e_T \) and \( e_L \) are the values of associative strength generaliz-
ing to the tone and light, respectively). The similarity of the tone
to the light–tone compound is given by \( TL_{ST} \), and the similarity
of the light to this compound is given by \( TL_{SL} \). For the sake of
clarification, the way in which the values of \( TL_{ST} \) and \( TL_{SL} \) can be
computed from Equation 2 will be described in some detail.
The stimuli common to the buffer on training trials with the
compound, and test trials with the tone, will be the tone and the
contextual stimuli, C. As a consequence, \( P_{con} \) will be deter-
mined by the value of the combined intensity of the tone and
the contextual stimuli, \( P_{TC} \). During training, the proportion of
the buffer allocated to the common stimuli will be determined by
the ratio \( P_{TC}/P_{TLC} \), where \( P_{TLC} \) represents the combined in-
tensity of the light, the tone, and the context. On the test trial,
the common elements will constitute the entire source of stimu-
lation and will therefore occupy the entire buffer. By substitut-
ing these terms in Equation 3, Equation 5 indicates the way in
which \( TL_{ST} \) is determined by the intensity of the various stim-
uli. The value of \( TL_{SL} \) can be derived in a similar manner:

\[ TL_{ST} = \frac{P_{TC} \cdot P_{TLC}}{P_{TC}} \] (5)

It should be evident that the relative intensities of T and L
will exert a profound influence on the values of S. In Equation
5, if the light is very much more intense than the tone, the value
of \( TL_{ST} \) will approach zero, because the value of \( P_{TLC} \) will be
very much greater than \( P_{TC} \). Equation 4a indicates in this in-
cidence, therefore, that the magnitude of generalization from the
compound to the tone will be slight. On the other hand, if the
intensity of the tone is very much greater than the light, the
value of \( P_{TC} \) will be only slightly less than \( P_{TLC} \). As a result,
\( TL_{ST} \) will approach one, and the strength of the CR elicited by
the tone will not be too different from that elicited by the com-
pound. In other words, after conditioning with the tone–light
compound there will be a loss of responding that may be re-
garded as generalization decrement whenever the tone or light
is presented separately. The magnitude of this decrement will
be greatest for the least intense stimulus. If the decrement
should result in the stimulus eliciting a weaker CR than after
conditioning with this CS alone, overshadowing could be said
to have occurred.

According to the view presented above, then, overshadowing
is due to generalization decrement, and its effects should be evi-
dent after one conditioning trial. It has already been noted that
there is ample support for this prediction that is not to be de-
ferred from the majority of contemporary theories of learning.

Application to external inhibition. External inhibition can be
readily explained by the model because adding a stimulus to a
CS will alter the contents of the buffer from those present during
conditioning. Thus, if subjects are trained with a light and then
tested with a light–tone compound, the value of \( TL_{ST} \) will deter-
mine the extent of generalization from the light to the com-
 pound. According to Equation 3, this parameter will be deter-
mined by the ratio of the intensity of stimulation when the light is
on to that when the light–tone compound is present.

There have been relatively few studies investigating external
inhibition, and it is appropriate, therefore, to cite the outcome
of several experiments reported by Young (1984). These were
designed to test two predictions from the model. The first pre-
diction is that there should be a symmetry between the effects of
overshadowing and external inhibition. Provided that the same
stimuli are employed, Equation 3 stipulates that the value of S
should be the same for both procedures. As a result, the general-
ization decrement engendered by the removal of an element
from a compound should be the same as that occurring when a
neutral stimulus is presented in compound with a previously
trained CS. To test this prediction, one group received condi-
tioning with a tone and extinction trials with a light–tone com-
pound. A second group received the opposite of this procedure:
compound conditioning followed by nonreinforced presenta-
tion of the tone. Hence, for both groups the transition to ex-
tinction was accompanied by the same magnitude of stimulus
change, produced either by addition or by removal of a light,
and this should result in a similar decrement in the strength
of conditioned responding. In support of this prediction, the
extinction curves for the two groups were very similar, with no
statistically significant difference between them. Moreover,
from the initial trial onward, the strength of the CR performed
by these subjects during extinction was significantly less than
for a group receiving the compound in both stages of the exper-
iment. This latter difference indicates that the stimulus change
in the first two groups was effective in producing a decrement
in the strength of responding.

The second prediction relates to the duration of the neutral
stimulus that is presented simultaneously with the pretrained
CS. Typically, studies of external inhibition have presented
these stimuli for the same duration, but according to the views
developed here this should not be necessary for a successful
demonstration of external inhibition. It also should be possible
to produce a decrement in the strength of conditioned respond-
ing by presenting a CS in conjunction with a stimulus that is on
for the entire test session. To test this proposal, two groups of
subjects received excitatory conditioning with a tone. For one
group the houselight was permanently on during this stage,
but the second group was trained in the dark. Both groups then
received a series of nonreinforced presentations of the tone with
the houselight permanently on. In keeping with expectations,
it was observed that extinction occurred more rapidly for the
group experiencing the change in background illumination at
the outset of the test stage.

Application to overexpectation. A number of authors have
noted that the results of overexpectation experiments are in-
compatible with analyses of Pavlovian conditioning similar to
that offered here (James & Wagner, 1980; Rescorla, 1970; Wagner,
1971). Specifically, in this procedure it is found that if two stimuli
that separately have been paired with a US are presented in
compound, the strength of the CR elicited by the
compound is greater than that elicited by either CS. In other
words, the presentation of a pretrained CS in compound with
another stimulus results in an effect opposite to generalization
decrement—namely, overexpectation. The present model does indeed predict a decrement in the strength of the CR generalizing to the compound from either element. But because the strength of the CR elicited by the compound will be determined by generalization from two sources, it is possible that the combined influence of these sources will exceed the original associative strength of either element. In essence, the expression $\lambda_S^{AB} \cdot E_A + \beta S^{AB} \cdot E_B$ will determine the strength of excitation generalizing to the compound from CSs A and B, and it is evident that overexpectation will occur whenever the combined values of the S parameter exceed 1, provided that conditioning has reached the same asymptote with each element. The combined values of the S parameters will always be greater than 1 in this type of overexpectation procedure, provided that the value of P for the contextual stimuli is greater than 0.

This discussion makes it evident that a number of sources of excitation can generalize to a CS. Equation 6 provides a general equation for determining the total excitation generalizing to CS A from n stimuli that are similar to A and have separately been paired with a US.

$$e_A = \sum_{j=1}^{n} /S_A \cdot E_j$$

(6)

Excitatory Conditioning

The discussion so far has been concerned with the generalization of excitation. In this section, the factors that determine whether or not a stimulus will acquire excitatory strength and to what extent are considered. In keeping with most theories of learning, repeated CS–US pairings are held to result in a negatively accelerated growth in excitatory associative strength. This view is expressed in Equation 7, in which $\Delta E_A$ is the increment in excitation strength on a given trial, and $\beta$ is a learning rate parameter with a value between 0 and 1 that is determined by the nature of the reinforcer.

$$\Delta E_A = \beta(\lambda - \bar{E}_A).$$

(7)

The term $\bar{E}_A$ is unusual, because it refers to the net or aggregate associative strength of the CS for that trial and can be formally represented by Equation 8:

$$\bar{E}_A = E_A + e_A.$$ 

(8)

In this equation, $e_A$ is determined in a manner according to Equation 6, and $E_A$ is the sum of the preceding values of $\Delta E_A$, that is, $E_A$ the current excitatory associative strength of CS. Thus, according to Equation 7, a change in associative strength will occur only when there is a discrepancy between the asymptote of conditioning and the combined excitatory strength of the CS. This in turn will be determined from its own prior pairings with the US and by generalization from other stimuli that have also been paired with the US.

Application to blocking. The fact that the excitatory strength of a CS can be derived from generalization from a similar CS that has already been paired with a US has important implications for the analysis of blocking. According to the model proposed here, pretraining with one CS, A, will result in it gaining asymptotic associative strength. When subjects are presented with a compound, AB, there will be a measure of generalization from A to AB. The discrepancy between this level of generalization strength and the asymptote of conditioning, $\lambda$, will permit some learning about the AB compound, but this will cease, according to Equation 7, when $E_{AB} + \lambda S^{AB} \cdot E_A = \lambda$. When, ultimately, B is presented by itself, there will be some generalization of the conditioned properties of the compound to this stimulus, of magnitude $AB^{SB} \cdot E_{AB}$. Although, according to this analysis, there may well be some evidence of a CR in the presence of B, it should be considerably weaker than that demonstrated by a standard control group that is not pretrained with A. This will be the case because compound conditioning will result in AB acquiring asymptotic associative strength and thus will permit a greater degree of generalization to B.

One prediction to follow from this analysis is that blocking will depend upon the relative intensity of A and B. If A, the pretrained element, is much less intense than B, $\lambda S^{AB}$ will be low, there will be very little generalization from A to AB, and as a result AB during compound conditioning will acquire a considerable degree of associative strength. Because $AB^{SB}$ will be relatively high, generalization from AB to B will also be considerable, and the results from such a study will not be too dissimilar to those obtained with an overshadowing procedure using the same stimuli. When the intensity of the pretrained element is much greater than the added stimulus, $\lambda S^{AB}$ will be high, and this will result in AB acquiring little associative strength on compound conditioning trials. Furthermore, the relatively low value of $AB^{SB}$ will result in a very weak CR being elicited by B when eventually it is presented in isolation. Although these predictions have not been tested rigorously in a single set of experiments, the little evidence that is available suggests they may well be confirmed. Hall, Mackintosh, Goodall, and Dal Martello (1977), for example, have reported that prior conditioning with a weak tone exerted only a slight blocking influence during compound conditioning when an intense light was the additional element.

In conclusion, it can be noted that blocking has served as a very important tool in the evaluation of a number of theories of learning. Mackintosh (1975), for instance, has justifiably placed considerable importance on his finding that blocking is not effective on the first trial of compound conditioning, but this result is not consistently observed (Balaz, Kaszrow, & Miller, 1982). The present analysis may not provide a comprehensive explanation for all the results obtained from studies of blocking. It does highlight, however, that one factor that may determine the outcome of blocking is the relative intensity of the elements. Until this influence is more fully explored it may be unwise to draw any firm theoretical conclusions from blocking studies.

Inhibitory Conditioning

Application to extinction. If, after excitatory conditioning, a CS is repeatedly presented by itself, a gradual weakening of the CR it elicits may be observed. At least two different processes have been proposed to account for this extinction effect. A number of theorists have suggested that the repeated independent presentations of a CS result in a gradual weakening of previously formed associations (e.g., Rescorla & Wagner, 1972). In contrast, theorists such as Konorski (1967; see also Pearce & Hall, 1980) maintain that extinction results in the formation of
new associations rather than the weakening of old ones. These new associations are assumed to be between internal representations of the CS and the absence of the US. To use Konorski’s (1967) terminology, extinction results in the formation of CS–no-US associations. Thus, when the CS is presented during extinction it will excite simultaneously activity in a US and a no-US center. Because it is assumed that activity in the no-US center inhibits activity in the US center and that the level of activity in the latter determines the strength of the CR, it follows that the stronger is the CS–no-US association the more pronounced will be the effects of extinction. For reasons elaborated elsewhere (Pearce & Hall, 1980), these ideas of Konorski (1967) will be adopted for the present model, using the term inhibitory learning to refer to the development of CS–no-US associations.

In order to describe the growth of these inhibitory associations, some further comment is needed concerning Equation 7. On a nonreinforced trial, the value of $\lambda$ will be 0, and as a result $\Delta E_A$ will have a negative value if the CS possesses net excitatory strength. Instead of resulting in a decrement in the current value of $E_A$, such a trial will result in inhibitory learning of magnitude $\Delta I_A$, which will be equal to the magnitude $|\Delta E_A|$ for that trial. The accumulation of these increments in inhibitory strength will determine the ultimate level of inhibition conditioned to the CS, $I_A$.

Just as excitation can generalize from one stimulus to another, so too it is assumed that inhibition will generalize according to the same principles. Thus if one stimulus, $A$, has acquired a measure of inhibitory strength, $I_A$, the presentation of a compound containing that stimulus, AB, will result in the generalization of inhibition of magnitude $A S_{AB} I_A$. The relationship expressed in Equation 2, therefore, can also be used to determine the magnitude of inhibition generalizing to CS A, $I_A$, from all other stimuli that have been involved in inhibitory learning. Specifically, this value will be determined by the expression $\Sigma A S_{AB} - I_A$. Because a stimulus can possess conditioned excitatory and inhibitory strength and generalized excitatory and inhibitory strength, its net associative strength, $\bar{V}_A$, will be determined by the difference between these levels of excitation and inhibition. Equation 9 expresses this relationship formally. Equation 10 shows how the value of $\bar{V}_A$, which determines the magnitude of the CR elicited by CS A, also determines the change in associative strength on a conditioning trial.

\[
\bar{V}_A = E_A + e_A - (I_A + I_A)
\]

\[
\Delta E_A = \beta (I_A - \bar{V}_A)
\]

During the course of straightforward extinction, $I_A$ will increase. This will produce a progressively smaller discrepancy between $E_A$ and $I_A$ and result in a correspondingly smaller difference between $\lambda$ (0) and $\bar{V}_A$ ($E_A - I_A$). Eventually a point of equilibrium will be reached when $E_A$ is equal to $I_A$. No further learning will occur, and no CR will be performed. The discussion now considers more complex procedures in which inhibitory learning is expected to occur.

**Application to A*AB discrimination.** One example of such a procedure is the case in which a stimulus possesses only generalized excitation and is nonreinforced. This will occur if subjects receive reinforced presentations of one stimulus, A*, intermixed among nonreinforced presentations of a compound involving that stimulus, AB*. If discrimination training is commenced once A has attained asymptotic excitatory strength, AB will elicit a CR of magnitude $A S_{AB} E_A$ because of the generalization of excitation from A. The nonreinforcement of AB will result in a growth of inhibition to the compound that will, to an extent, generalize to A. According to Equation 9 this generalized inhibition will detract from the net associative strength of A and will permit further excitatory learning with this stimulus. Eventually, the inhibitory strength associated with AB will nullify the excitation generalizing to this stimulus from A. In addition, the excitatory strength of A minus the inhibitory strength generalizing to it from AB will equal $\lambda$. At this point, the CR elicited by A will be the same as for subjects receiving only excitatory conditioning with A. There will also be no evidence of a CR in the presence of AB.

Figure 1 indicates the trial-by-trial changes in associative strength that are predicted by Equation 10 to occur during discrimination training of the form A*AB*.
from A to B, and B will thus possess a degree of net inhibitory strength, which can be seen to be the case in the lower part of Figure 1. This, according to Equation 10, will retard excitatory conditioning with B relative to a control group trained with a novel CS.

As far as the summation test is concerned, assume that the influence of B is being assessed with a conditioned excitor C. The strength of the CR elicited by the compound, BC, will be weaker than that elicited by C for two reasons. First, this will be because of the generalization decrement induced by the presence of B, irrespective of its previous discrimination training. Second, detracting further from this value will be the inhibition generalizing from AB to BC, which will be of magnitude \( A_B S_{BC} - I_{AB} \). Assuming roughly similar intensities of A, B and C, the parameter \( A_B S_{BC} \) will be less than \( C_S_{BC} \). As a result, the generalization of excitation from C to BC will exceed the generalization of inhibition from AB, and the BC compound will possess net excitatory strength. In contrast, it has already been shown that as a result of straightforward A'AB' discrimination training the AB compound will eventually possess zero net associative strength. Thus, the model predicts that a putative conditioned inhibitor will be more effective in a summation test when it is presented in compound with a stimulus with which it was originally trained than when it is presented in conjunction with a stimulus with which it has never previously been paired. It is of some encouragement to note that this result is generally observed (e.g., Rescorla, 1982). Moreover, this finding is inconsistent with theories that assume that the conditioned properties of a CS will be unaffected by changing the context in which it is tested from that in which it was trained.

**Application to an A'AB' discrimination.** In a feature-positive procedure, subjects receive training in which one stimulus is nonreinforced, A', and a compound comprising this stimulus and another element is reinforced, AB'. As a result, a conditioned response will eventually occur whenever the compound but not the element is presented. According to the present account, mastery of the discrimination will be due to the following process. On reinforced trials the compound will acquire excitatory strength that will generalize to A. The nonreinforcement of this stimulus will result in it acquiring inhibitory strength to counteract the generalized excitation, and this inhibition will in turn generalize to the compound and necessitate a further growth of excitatory strength to A. At asymptote, therefore, the magnitude of inhibition conditioned to A will be equal to the excitation generalizing to it from AB, and no CR will be observed in the presence of this stimulus. The inhibition generalizing to AB from A will subtract from the now supernormal level of conditioning with the compound and ensure that its net associative strength is the same as if only reinforced compound trials had been administered.

Figure 2 shows the trial-by-trial changes in the excitatory and inhibitory strengths of A and AB predicted by Equation 10 to occur during the course of the discrimination. The values of the various parameters were the same as for the example in the preceding section.

In discussing the Rescorla-Wagner model it was noted that for the feature-positive procedure it predicts that once the discrimination has reached asymptote, the strength of the CR elicited by the compound will be determined entirely by the associative strength of the feature CS, B. This prediction is also made by other contemporary theories of conditioning (e.g., Mackintosh, 1975; Pearce & Hall, 1980). It follows from these theories, therefore, that presenting B by itself will elicit a CR of comparable magnitude to that elicited by AB. In contrast, the present analysis predicts that the strength of the CR elicited by the feature stimulus, B, alone will be determined by its relative intensity to A. If B is much more intense than A, the value \( A_B S_B \) will be high, and the CR generalizing to B from AB will be comparatively strong. On the other hand, if the intensity of B is low relative to A, \( A_B S_B \) will be low, and a much weaker CR will generalize to B from AB. To test these predictions, Young and Pearce (1984) examined the effects of feature-positive training with a light and tone. In order to assess the relative intensities of these stimuli, two groups received straightforward compound conditioning. When they were subsequently presented in isolation, it was found that the tone elicited a considerably stronger CR than the light, implying that the tone was a more intense stimulus than the light and that the value of \( T_L S_L \) was considerably less than \( T_L S_T \). Accordingly, after a feature-positive discrimination it can be predicted that when the light is employed as the feature stimulus it will elicit a weaker CR when tested in isolation than if the tone constitutes the feature element. This was indeed the case, and it is worth remarking that we took steps to ensure that the discrimination had reached asymptote for both cases. Thus, for example, the discrepancy in the conditioned properties of the light and the tone was not at all diminished by increasing the discrimination training by as much as 50%.

When the model was introduced it was stated that on any conditioning trial the representation in the buffer will partly be determined by the context in which conditioning takes place. As a result, the analysis of the acquisition of associative strength to a CS reinforced in isolation is formally equivalent to that of
the feature-positive procedure just described. That is, subjects will be required to discriminate between a representation of the CS plus the context and a representation of the context alone. One implication of this equivalence is that the salience of a CS will determine its conditionability—that is, the rate at which it will acquire associative strength.

By definition, the intensity of a CS is held to determine the proportion of the buffer that it occupies whenever it is presented. Because the buffer is held to be of fixed capacity, it follows that contextual stimuli will be granted less space when they accompany a CS of high rather than low intensity. Thus, when conditioning is conducted with a very intense CS, there will be only a slight generalization of excitatory strength of the contextual stimuli present during the intertrial interval. This will result in the CR being confined to the CS from virtually the outset of training. However, the slight generalization of excitation to the context (providing that reinforcement is withheld during the intertrial interval) will result in the context acquiring a measure of inhibition. The generalization of this inhibition to the CS will counteract to a slight extent the effects of the prior excitatory training. Eventually a state of equilibrium will be arrived at in which the inhibitory strength of the context will compensate for excitation generalizing to it from the CS, and supernormal conditioning with the CS will compensate for the inhibition generalizing to it from the context. In the case of training with a less salient CS there will be a greater degree of generalization from the CS to the context. As a result, the level of inhibition associated with the context will be correspondingly larger, and its generalization to the CS will exert a considerably disruptive influence on excitatory conditioning. Thus, the greater the intensity of a CS the more rapidly will it acquire an overt CR and the higher will be its conditionability.

Application to relative validity. The discussion of feature-positive training leads to a consideration of the importance of "relative validity" in determining the outcome of compound conditioning. Wagner, Logan, Haberlandt, and Price (1968) trained one group in a pseudodiscrimination of the form AB+0 AC+0, in which the simultaneous compounds AB and AC were reinforced on 50% of their occurrences. A second group was trained with a true discrimination of the form AB AC0, in which AB was consistently reinforced and AC was consistently nonreinforced. In both groups A was equally often paired with the US, yet this stimulus acquired greater associative strength in the first group, where it was as valid as B and C as a predictor of the US, than in the second group, where its predictive validity was less than that of B and C. On the basis of this type of result, Wagner et al. (1968) concluded that the relative validity of the elements of a compound as predictors of reinforcement is an important determinant of their ultimate associative strength. Wagner (1971) has described a similar outcome using the feature-positive procedure. Thus, the strength of the CR elicited by the feature element, a more valid predictor of the US than the CS with which it was paired during conditioning, was greater than that elicited by the equivalent stimulus in an overshadowing group, in which both stimuli were equally valid as predictors of the US.

At first sight it might appear that these results conflict with the tenor of the model being proposed. When subjects are presented with an element, after compound training, the strength of the CR it elicits is not determined solely by the degree of stimulus change but also by the relative validity of the elements of the compound. This outcome, however, can be seen to follow from the model once it is acknowledged that nonreinforcement results in inhibitory conditioning that may generalize. Turning first to the feature-positive procedure, it was noted above that as a result of this training, the level of excitatory conditioning with the compound will exceed that obtained in a straightforward overshadowing design. In essence, this supernormal conditioning will be due to, and masked by, inhibition generalizing to AB from A. When B is subsequently presented by itself, therefore, there will be more scope for the generalization of excitation from AB as a result of the feature-positive training, and a stronger CR will be expected than with overshadowing. Moreover, the difference between A and B will mean that there will be very little generalization of inhibition to oppose this effect. The analysis of the results reported by Wagner et al. (1968) is potentially complex but can be simplified once it is noted that the strength of the CR elicited by the partially reinforced compounds in the pseudodiscrimination group was only slightly less than that elicited by AB in the true discrimination procedure. The implication of this outcome is that the associative strength of the partially reinforced compounds was not much less than for the consistently reinforced compound. Partial reinforcement with both AB and AC will therefore provide two sources of excitation to generalize to A, and their combined influence will result in a relatively strong CR to this stimulus for those subjects trained with the pseudodiscrimination. But for the group receiving the true discrimination, there will be only one source of excitation generalizing to A, that due to the continuously reinforced AB compound. Moreover, the effect of this generalized excitation will be diminished by the generalization of inhibition from the consistent nonreinforcement with AC.

Development of the Model

The main concern of the preceding section has been to introduce the model in detail and to demonstrate how it can be applied in a variety of conditioning procedures. In this section an attempt is made to show how the model can be extended to deal with some more general issues of conditioning.

The Influence of Memory and Attention on the Buffer

Hitherto the extent to which a stimulus is represented in the buffer has been determined by its intensity relative to the intensity of all the stimuli present on a trial. It is possible that processes such as memory and attention can influence the contents of the buffer. The implications of this possibility are considered briefly in this section.

With respect to memory, it is reasonable to assume that once a stimulus is removed from the environment its representation may persist for a while in the buffer. Presumably the area allocated to the stimulus will be less than for the stimulus itself, but currently little more can be said about the factors that influence the durability of a memory in the buffer. The persistence of such a trace, however, would account for effects such as trace conditioning and conditional discriminations involving the serial presentation of stimuli.
Considerably more can be said about the way in which selective attention might influence the contents of the buffer. Conceivably, a process of this sort could modify the salience or effective intensity of a stimulus prior to its entry into the buffer. For example, it has been observed repeatedly (e.g., Sokolov, 1963) that the presentation of a novel stimulus results in an orienting response that wanes with its repeated presentation. Such a response is considered to enable subjects to attend maximally to a stimulus, and this may result in it initially gaining access to a considerable proportion of the buffer. As the orienting response declines in strength, however, the proportion of the buffer occupied by the stimulus might also be expected to decline. Because the area in the buffer occupied by a stimulus determines its conditionability, it follows that subjects should condition more readily with a novel than a familiar stimulus. The results from experimental investigations of latent inhibition provide ample support for this prediction (cf. Lubow & Moore, 1959).

It is also possible that during compound conditioning subjects are predisposed to attend to one of the elements to a greater extent than the other and that this bias is determined by the nature of the US. As a result, the extent of generalization from the compound to the separate elements may differ for aversive and appetitive training procedures. In support of this proposal it can be noted that pairing an audiovisual compound with a shock US results in the auditory element exerting a greater control over responding than the visual element when they are tested independently. This pattern of results is reversed when appetitive conditioning is conducted with the same compound (Lolordo, 1979; Lolordo & Furrow, 1976; Schindler & Weiss, 1982).

The Mediation of Generalization

A fundamental feature of the model as it has been presented so far is that the effects of conditioning can generalize only to stimulus configurations that are physically similar to the initially trained CS. Hence, if subjects are trained with A and are then presented with AB, the compound will elicit a CR because of its similarity to A. Although B is similar to AB, it has been maintained that this stimulus will not elicit a CR, if the role of the context is ignored, because it bears no similarity to A. One way of characterizing this account is to suggest that whenever a CS is represented in the buffer it activates similar representations in LTM. The extent of this activation could be determined by the value of S relating the contents of the buffer to the LTM representation. The strength of the CR elicited by the representation in LTM should then be a joint function of its extent of activation, S, and its associative strength, E - I. As it is presently formulated, the model asserts that an active representation in LTM is incapable itself of activating similar representations. In fact, this restriction may be incorrect.

It is quite conceivable that once a representation is active in LTM, it will activate other representations that are similar to it. The extent of this mediated generalization might be determined by the product of the arousal of the initially activated representation and the value of S relating it to the similar representation in LTM. Such a relation would concentrate considerably the spread of activation through LTM. Thus if a CS, A, is represented in the buffer, it will activate a representation of AB in LTM to a magnitude $a_{AB}$. The activation of AB will in turn excite the LTM representation of B to the extent $a_{AB}B$. According to this modification, therefore, the presence of one stimulus in the buffer may be able to excite a representation of a CS that is totally dissimilar to it, but in many instances the extent of this activation will be slight.

To demonstrate the implications of this approach it is useful to consider a blocking experiment in which conditioning with A is conducted prior to compound training with AB. Subsequent presentations of B will then excite a representation of AB and to a lesser extent A. The activation of these separate representations will each excite a CR to different degrees, but their combined influence need not be particularly large. If, after conditioning with AB, subjects receive nonreinforced presentations of A, the subsequent CR elicited by B will be weakened because although the representation of A will be activated, it will no longer be capable of contributing to the CR. This prediction has been confirmed in a series of experiments reported by Rescorla and Durlach (1981, p. 91). Indeed, the present analysis bears many similarities to Rescorla's account of within-event learning or simultaneous associations (Rescorla, 1981; Rescorla & Durlach, 1981), as the following discussion indicates.

Hitherto it has been proposed that the contents of the buffer are stored in LTM only when they are involved in excitatory or inhibitory conditioning. The results from sensory preconditioning experiments indicate that stimuli not involved in conditioning procedures may also be represented in LTM. For example, Rescorla and Cunningham (1978) exposed subjects to a simultaneous compound composed of two tastes prior to pairing one of the tastes with an aversive US. As a result of this training, the other element of the compound elicited a CR when subsequently it was presented alone. In terms of the present formulation, the nonreinforced compound trials would have enabled the formation of its representation in LTM. This representation could then mediate the generalization of excitatory conditioning from the conditioned element to the nonreinforced test element in a manner analogous to that described above for blocking.

Unfortunately, it is not possible to make any precise specifications concerning the factors that promote the formation of nonreinforced compound stimulus representations in LTM. But it is possible to identify a number of variables that might enhance the mediation of generalization by such representations once they are formed. First, if representations of AB and B already exist in LTM, the presence of A in the buffer will activate B to a magnitude $a_{AB}AB$. According to the previous discussion, the larger the value of this product the greater will be the influence of A in exciting a CR due to B's activation. A method for ensuring that both $S$ values are relatively high is to select stimuli that will be equally represented in the buffer and to a large extent. If this is not achieved, the magnitude of one of the $S$ parameters will be low, and their combined value will also be low. The stimuli should thus be of equivalently high intensity and preferably novel, because nonreinforced exposure to either of them might diminish the extent to which they are represented in the buffer when presented in compound. Second, if A and B are themselves similar, both $S$ values will be relatively high because each stimulus will have elements in common with
its associate in the AB representation. Thus, even if discrimination training is administered to minimize generalization between A and B directly, the presence of an AB representation in LTM will permit the mediation of considerable generalization between those stimuli. Third, if the AB compound is presented serially with A preceding B, the representation of the compound will comprise B plus the trace of A, which, if it has already been noted, might reduce the effective intensity of A. As a result the value of \( S_{AB} \) will be slight and certainly less than if A and B were presented simultaneously. The mediation of generalization can thus be expected to be most effective by compounds that are simultaneously rather than serially presented. Finally, the stimuli comprising the compound should be presented in a similar location, or in a manner that will ensure they are perceived simultaneously. If this is not achieved, it is unlikely that A and B will be represented simultaneously in the buffer on compound trials.

The foregoing proposals are derived from assumptions about the operation of the buffer and the rules governing the magnitude of generalization. It is therefore particularly encouraging to note that much of what has just been suggested is corroborated by the outcome of experiments investigating within-event learning. For example, Rescorla (1980) has demonstrated that sensory preconditioning is more effective when the stimuli are presented simultaneously rather than serially. Rescorla and Furrow (1977) have demonstrated that the use of similar rather than dissimilar stimuli results in better second-order conditioning, which implies the development of a better mediator.

**An Alternative Formulation**

The above model assumes that the representation of a CS consists of the entire pattern of stimulation on a trial, and in this respect it can be considered to be a configural theory of learning. It is possible, however, to overcome many of the problems identified in the first section of this article by adopting an alternative, more elemental view of conditioning.

It is conceivable that any single stimulus is composed of a number of elements, some of which are removed whenever it is presented in compound with another stimulus. Thus if one group receives a compound conditioning trial with AB, whereas another is given a single trial with A alone, the subsequent presentation of A will result in a weaker CR being performed by the former group. This follows because more elements of A will be available for conditioning when it is presented alone than on the compound trial. It should be evident that a similar analysis to this account of one-trial overshadowing can also account for external inhibition, for the feature-positive results that were described earlier, and for configural learning such as that involved in patterning schedules.

Attractive as this proposal may be for its simplicity, it has not been developed further in this article, because so little is known about the properties of the elements of which a stimulus might be composed. In particular, until more is known about the way they are affected by the context in which they occur, it will be difficult to derive precise predictions from this perspective.

**References**


Received October 22, 1984

Revision received April 2, 1986