Psychological Review
1980, Vol. 87, No. 6, 532-552

A Model for Pavlovian Learning: Variations in the Effectiveness of Conditioned But Not of Unconditioned Stimuli

John M. Pearce
University of Cambridge, Cambridge, England

Geoffrey Hall
University of York, York, England

Several formal models of excitatory classical conditioning are reviewed. It is suggested that a central problem for all of them is the explanation of cases in which learning does not occur in spite of the fact that the conditioned stimulus is a signal for the reinforcer. We propose a new model that deals with this problem by specifying that certain procedures cause a conditioned stimulus (CS) to lose effectiveness; in particular, we argue that a CS will lose associability when its consequences are accurately predicted. In contrast to other current models, the effectiveness of the reinforcer remains constant throughout conditioning. The second part of the article presents a reformulation of the nature of the learning produced by inhibitory-conditioning procedures and a discussion of the way in which such learning can be accommodated within the model outlined for excitatory learning.

A recent review of classical conditioning (Dickinson & Mackintosh, 1978) suggests that the underlying associative process is "responsible for organisms learning about the relationships between events, enabling them to build up an associative representation of the causal structure of their environment" (p. 588). According to this view the occurrence of a conditioned response (CR) as a result of the pairing of a conditioned stimulus (CS) and an unconditioned stimulus (US) is regarded as being merely an index of the formation of some internal representation of the relationship between CS and US. The first task of a theory of classical conditioning becomes that of specifying exactly the nature of this internal representation, and in recent years there has been fairly general agreement about the nature of the representation formed as a result of simple conditioning procedures. It is usually assumed that an association is formed between the central representations of the CS and US so that activation of the first (by presentation of the CS) arouses activity appropriate to the likely occurrence of the second (e.g., Bindra, 1972; Konorski, 1948; Mackintosh, 1974). In consequence, the "associative strength" of the CS has become a central concept in classical conditioning theory, and the concern of the theorist has been largely to specify how various procedural manipulations work to determine this strength (e.g., Frey & Sears, 1978; Mackintosh, 1975b; Rescorla & Wagner, 1972).

Theories of this sort have faced two main problems. The first arises from the observation that in a number of circumstances, the pairing of CS and US does not result in learning. Of the various theoretical mechanisms proposed to deal with this fact, none can encompass all of the data. The first major section of this article reviews the ways in

This work was supported by grants from the United Kingdom Science and Medical Research Councils.

We have made free use of many suggestions, especially concerning inhibitory learning, made by A. Dickinson. We thank him and also C. Adams, P. Bailey, S. Channell, N. Mackintosh, and E. Macphail for their comments on earlier versions of this article.

Requests for reprints should be sent to John Pearce, Department of Psychology, University College, Cardiff, U.K. CF1 1XL or Geoffrey Hall, Department of Psychology, University of York, York, England Y01 5DD.

which the theories fail and presents as a formal model an alternative that we hope can deal with more of the data than its competitors.

The second problem arises from the observation that learning of a rather special sort (inhibitory learning) occurs when the CS and the US are *not* paired. The attempt by Wagner and Rescorla (1972) to apply their basic model for excitatory conditioning to the inhibitory case yielded some notable successes but ran into a number of difficulties (see Rescorla, 1979, for a review) that still await a fully satisfactory explanation. We believe that these difficulties arise from the failure of current models for inhibitory conditioning to specify appropriately the nature of the internal representation produced by these procedures. Accordingly, the second major section of this article begins with an attempt to clarify this issue. Having achieved this reformulation, we hope to show that inhibitory learning can be readily accommodated by the model that we propose for excitation. Indeed, one aim of this article is to argue that there is no essential difference between the learning produced by inhibitory and excitatory conditioning procedures.

Excitatory Learning

The Problem

We begin by making the assumption that when an animal experiences an appropriate CS and US in close contiguity, their conjoint processing results in the strengthening of an association between their internal representations. We shall symbolize the strength of this association as V; we take the size of the increase in V (ΔV) produced by such a processing episode to be some function of the strength or effectiveness of the internal representations of the CS and US. The simplest way of expressing these factors that can serve as a starting point for discussion is given as Equation 1:

$$\Delta V = \alpha\lambda, \qquad (1)$$

where α and λ are parameters varying with CS and US intensity, respectively. (The exact functions relating α and λ to physically measured stimulus intensity will depend on details of the animal's sensory system.) There is no doubt that the speed and completeness of conditioning do depend on the physical intensity of the stimuli, but there is equally no doubt that other factors play a part in determining the effectiveness of the CS or the US or both.

The most obvious demonstration of the fact that the pairing of a seemingly adequate CS and US can fail to produce an increase in associative strength is seen when conditioning reaches an asymptote. Clearly, as conditioning proceeds, something happens that causes ΔV to decline. A more subtle example of the same basic phenomenon is supplied by the occurrence of "blocking" (Kamin, 1969). Here, the pairing of a CS and US, which have been separately demonstrated to be capable of supporting conditioning, produces no increase in the associative strength of the CS; in some way the simultaneous presence of another stimulus that has already formed an association with the US "blocks" the acquisition of associative strength by the CS in which we are interested.

Models of classical conditioning have explained these failures to learn by suggesting that conditioning procedures can cause stimuli to lose their effectiveness. Dickinson and Mackintosh (1978) have suggested that such models can be divided into two main categories: those that suggest that the US may become ineffective and those suggesting that the CS may become ineffective.

Variations in US Effectiveness

Kamin's own explanation for blocking embodied the suggestion that a US that has already entered into a full association with a CS is not effective when that CS is present. The Rescorla-Wagner (1972) model expresses some aspects of Kamin's ideas in a formal way. It proposes that the change in associative strength of Stimulus A depends in part on the discrepancy between the asymptote of conditioning (determined by the intensity of the US, λ) and the sum

of the associative strengths of all CSs present (V_Σ).[1]

$$\Delta V_A = \alpha(\lambda - V_\Sigma). \quad (2)$$

Thus, the effectiveness of the US is given by the expression $(\lambda - V_\Sigma)$, and when V_Σ is equal to λ, the US will be completely ineffective. Blocking will occur; that is, there will be no increase in V_A when some other CS is simultaneously present that has already been conditioned so that $V_\Sigma = \lambda$. It should be apparent also that the increments in the associative strength of a CS conditioned in isolation should decline to zero as its own increasing associative strength reduces the effectiveness of the US. This theoretical position has often been summarized as saying that a US is only effective when it is not predicted or is "surprising."

There is a sense in which this theory is almost too successful in its ability to predict blocking; we discuss next two cases in which no blocking occurs when the model predicts that it should. The first can be taken to show that although the surprisingness of the reinforcer is an important determinant of learning, it exerts its effect in a way different from that proposed by Rescorla and Wagner. Dickinson, Hall, and Mackintosh (1976) carried out a blocking experiment in which rats were pretrained with a stimulus, B, the US being two electric shocks separated by an interval of 8 sec. One group then received a compound stimulus, AB, followed by this same reinforcer, and these subjects showed blocking; that is, they failed to learn about Stimulus A. A second group received compound training in which the second of the two shocks was omitted; for these animals, Stimulus A did acquire associative strength. Dickinson et al. concluded that the surprising omission of the second shock can cause "unblocking." It should be apparent from Equation 2 that according to the Rescorla-Wagner model, unblocking should occur only when λ is increased or V_Σ is reduced on the compound trials. Neither of these conditions was met in the experiment by Dickinson et al.

The other case challenges what is perhaps a more central feature of the Rescorla-Wagner account. Mackintosh (1975a) carried out a blocking experiment in which, after Stimulus B had been established as a CS for shock, only a single AB trial was given. He found that Stimulus A acquired a small but measurable amount of associative strength as a result of this trial. The performance of other groups of subjects served to show that further reinforced trials did not produce any further increase in strength to A (i.e., that blocking did occur after the first trial), also that the strength gained by A was the same as that produced by a single compound trial not preceded by pretraining with Stimulus B (i.e., that conditioning was normal on the first trial). Mackintosh's own explanation of this finding will be discussed later. For our present purpose it is sufficient to point out that this result undermines the central assumption of models of the Rescorla-Wagner type. For here is a case in which a fully predicted and unsurprising US turns out to be a perfectly effective reinforcer. At the very least, therefore, we must allow that some additional mechanism is involved in blocking and related phenomena. We turn next to a consideration of two theories that suppose there to be some change in CS effectiveness as a result of conditioning in addition to any loss of US effectiveness.

Before doing so it is only fair to add at this point that Wagner and Rescorla (1972) allowed that the effectiveness of a CS (i.e., α) might change. To explain "latent inhibition," the retardation of conditioning produced by prior presentation of the CS alone, they added the assumption that such treatment might cause a decline in α for that stimulus. This is an important step, since it admits the possibility that the effectiveness of a CS may be determined by factors other than the physical intensity of the stimulus. But this idea was taken no further; changes in α were used only to explain latent inhibition and played no part in the model's

[1] This model (and similar models) also includes a learning-rate parameter dependent on the properties of the US. For clarity of exposition, this parameter has been neglected throughout; its inclusion does not materially affect the application of these models to the issues discussed in this article.

account of blocking, and so on, nor was there any formal statement of the rules governing changes in α.

Variations in CS Effectiveness

The Rescorla-Wagner model supposes that a US loses its effectiveness as it comes to be accurately predicted by preceding events. It has proved possible to develop similar theory to deal with changes in CS effectiveness, the central assumption being that a CS, too, can become associated with events that predict it and thus become unsurprising. We consider an example of this sort of theory first but then go on to describe an alternative that makes the contrary assumption that the effectiveness of a CS is determined not by the extent to which it is predicted but by the extent to which it predicts its consequences.

CS Effectiveness and the Predictability of the CS

Wagner (1978) has put forward an elaboration of the Rescorla-Wagner (1972) model, expressed in rather different terms. He retains the premise that the surprisingness of the US as given by $(\lambda - V_\Sigma)$ determines how much will be learned on any given trial, but a version of this premise is extended to govern what we may call the "associability" of the CS. Wagner maintains that the presentation of a CS in a given context will result in the formation of an association between the CS and that context, and it is the strength of this association that determines the surprisingness of the CS and hence its associability. Just as a fully expected US loses its ability to reinforce conditioning, so a fully expected CS loses its ability to become associated with a US. More formally, we may say that the parameter α may change as a result of training, and if l is the asymptotic strength of the context-CS association and v is the current strength of this association, then α is given by Equation 3 as

$$\alpha = \delta(l - v) \tag{3}$$

where δ is a learning-rate parameter determined by the properties of the CS, which takes a value between 0 and 1. By substituting for α in Equation 2, we derive

$$\Delta V_A = \delta(l - v)(\lambda - V_\Sigma), \tag{4}$$

which describes how changes in α can modify the course of conditioning.[2]

The application of this model to the loss of associability seen in simple latent inhibition is straightforward. The repeated presentation of a stimulus prior to conditioning will permit the development of a strong context-stimulus association, and the value of the term $(l - v)$ will tend to decline. It follows that the acquisition of associative strength by this stimulus when it is presented with a reinforcer will proceed only slowly. This basic account can be extended to deal with more complex cases of latent inhibition. For instance, it has been established that the loss of associability produced by nonreinforced preexposure of a stimulus is attenuated when that stimulus is followed by some other event (Hall & Pearce, 1979; Lubow, Schnur, & Rifkin, 1976). Wagner's suggestion is that the processing of stimuli that results in the formation of an association between them takes place in a device of limited capacity. The occurrence of some event, particularly an unexpected one, following the presentation of two stimuli will take up some processing capacity, thus reducing the capacity that can be devoted to the stimuli in which we are interested, in this case producing a weaker than usual context-CS association.

From the little we have said, it should already be apparent that Wagner's model has greater explanatory power than the Rescorla-Wagner model on which it is based. Wagner (1978) described the successful application of his model to a wide range of data that lie outside the scope of any of the other models discussed in this article.

---

[2] The model presented by Wagner (1978) is in fact more elaborate than that represented by Equation 4. In particular, Wagner argues that in addition to the conditioned changes in associability depicted here, there are also transient changes in associability produced by the presentation of the CS. This has proved to be a fruitful idea, but it will not be pursued further here, since all experiments considered here are concerned with longer term changes in associability.

Table 1
Summary of Procedure Used by Mackintosh, Bygrave, and Picton (1977)

Group	Stage 1 (4 trials)	Stage 2	
		Trial 1	Trial 2
+	L+	TL+	—
++	L+	TL++	—
+/+	L+	TL+	TL+
+/++	L+	TL+	TL++
++/+	L+	TL++	TL+
++/++	L+	TL++	TL++

Note. L = light, T = tone, + = single shock, ++ = two shocks 10 sec apart.

Nonetheless, although the model deals in some detail with changes in CS effectiveness, it still retains the assumption that changes occur in US effectiveness according to how well the reinforcer is predicted, that the increment in associative strength on any trial is determined by $(\lambda - V_\Sigma)$. Consequently, this model is no better equipped than its predecessor to deal with the fact that blocking effects are not evident after just one trial; nor is it better equipped to explain the "unblocking" results of Dickinson et al. (1976), particularly when these are a consequence of shock omission. We turn now to a theory explicitly designed to accommodate these facts.

CS Effectiveness and the Predictive Power of the CS

Mackintosh (1975b) held that the associability of a stimulus is determined by how well it predicts its consequences. He suggested that the associability (α) of a stimulus will increase if it predicts reinforcement more accurately than other stimuli present in the situation but will decrease if it predicts reinforcement less accurately. These suggestions are formally expressed in Equations 5 and 6 taken from Mackintosh (1975b).

$$\Delta\alpha_A \text{ is positive if } |\lambda - V_A| < |\lambda - V_X|, \quad (5)$$
$$\Delta\alpha_A \text{ is negative if } |\lambda - V_A| \geq |\lambda - V_X|, \quad (6)$$

where V_X represents the associative strength of all stimuli other than A present on a given trial.

The manner in which α_A influences the acquisition of associative strength by A is expressed in Equation 7.

$$\Delta V_A = \alpha_A(\lambda - V_A). \quad (7)$$

Thus, Mackintosh rejects the view adopted by Rescorla and Wagner that change in the associative strength of a stimulus is determined by the discrepancy between λ and the sum of associative strengths of all the stimuli that happen to be present (V_Σ). Instead, the size of the associative change is determined by the discrepancy between λ and the strength of the stimulus in question. A result of this view is that the loss of US effectiveness is relegated to a fairly minor role in explaining conditioning effects; indeed, it now serves the role simply of explaining why the acquisition of associative strength by a given stimulus should eventually reach an asymptote fixed by the strength of the US. In blocking and related phenomena, the whole of the explanatory burden is borne by the mechanisms dealing with changes in CS effectiveness.

This model is applied to blocking as follows: It is assumed that α for a stimulus experienced for the first time is determined solely by its intensity. Thus, when a stimulus, A, is added to a pretrained element, B, in a blocking experiment, conditioning proceeds normally according to Equation 7, and there is an increment in the associative strength of A. After a conditioning episode, however, the α values of the stimuli experienced are modified according to the associative strengths that the stimuli have acquired as a result of that episode (i.e., according to Equations 5 and 6). After the first compound trial of a blocking experiment, Stimulus A will have acquired some associative strength, but the pretrained element, B, will be at or near asymptote. Thus $|\lambda - V_A|$ will be greater than $|\lambda - V_B|$, and α_A will decline. This means that the model explains why Stimulus A conditions normally on the first compound trial but hardly at all on subsequent trials.

The application of this model to the unblocking results of Dickinson et al. (1976) is dealt with most conveniently by considering first a related experiment by Mackintosh, Bygrave, and Picton (1977). This experiment is discussed in some detail, since it provides

evidence directly relevant to what we take to be the model's most important innovation, that is, its suggestion that the outcome of one trial can be used to set the associability of a stimulus and thus determine the course of conditioning on the next trial.

An outline of the experiment by Mackintosh et al. (1977) is given in Table 1. Six groups of rats received four trials on which a light was followed by an electric shock. They then received training with a tone–light compound. The groups differed in the number of compound trials they received and in whether or not a surprising second shock occurred 10 sec after the first. It was found that some conditioning occurred to the tone in all groups, with the smallest being found in the pair of groups that received only one compound trial. These two groups did not differ from one another; that is, the surprising second shock in Group ++ produced no more conditioning than the single shock given to Group +. Thus, the extra shock did not influence conditioning on the trial on which it was presented. A comparison of the remaining groups that received two compound trials reveals, however, that the surprising second shock on the first trial was certainly not without effect. The groups receiving a surprising shock on the first compound trial (Group ++/+ and Group ++/++) did not themselves differ, but they showed significantly more conditioning to the tone than the groups receiving a single shock on the first compound trial (Group +/+ and Group +/++). It seems, therefore, that a surprising shock on the first compound trial potentiates conditioning on the next trial. According to Mackintosh, the tone conditions normally on the first compound trial in all groups but loses associability when a single shock occurs on this trial, since it predicts this outcome less well than the light. The tone therefore acquires little or no further associative strength. But when a surprising second shock occurs, this loss of associability is greatly attenuated, since the light no longer accurately predicts the outcome of the trial. The tone is therefore able to acquire further strength on the second compound trial. An argument similar in principle (but rather more complex, see Dickinson & Mackintosh, 1979) can be applied to the unblocking effect produced by surprising shock omission (Dickinson et al., 1976).

The success of Mackintosh's model in explaining blocking and related phenomena convinces us that the principle it embodies—the modification of CS associability as a result of the consequences of one trial influencing conditioning on the next—must be a part of any successful theory. But there is reason to doubt the adequacy of the rules that Mackintosh suggests should govern changes in α. The inadequacy becomes apparent when we consider the application of this model to latent inhibition.

The decline in associability of a stimulus presented in isolation causes no problems and is governed by Equation 6. During latent inhibition training, λ, V_A, and V_X are all zero and thus α will decline. But in a recent experiment, Hall and Pearce (1979; see also Pearce & Hall, 1979) found that a similar decline in associability could occur even when the stimulus was preexposed not in isolation but in conjunction with a reinforcer. During the preexposure phase, subjects in the experimental condition received conditioning trials in which a stimulus (such as a tone) was paired with a weak electric shock. Control subjects received similar treatment with a different stimulus, such as a light. In the test stage the tone was used to signal a stronger shock for both groups. It was found that learning occurred relatively slowly in the experimental group (in spite of the fact that the stimulus in question had already acquired some associative strength as a result of the first stage of training).[3] Apparently, a stimulus can lose associability even as it acquires associative strength, but this phenomenon is entirely ruled out by Mackintosh's formulation. The tone in the experiment by Hall and Pearce is a better predictor of the reinforcer than any other stimulus, and hence, given

[3] The experiment by Hall and Pearce (1979) used the suppression of lever pressing as the index of conditioning. It is encouraging to note that Bolles and Sigmundi (Note 1) have recently reported an identical pattern of results in an experiment in which the extent to which the subjects were immobile during an aversive CS was used as the measure of conditioning.

Equation 5, its associability should go up during the preexposure phase. Thus, although we are willing to accept Mackintosh's suggestion that changes in CS effectiveness make an important contribution to the outcome of many conditioning procedures, we cannot accept the rules that he suggests might govern how these changes are brought about.

A New Account of CS Processing

The model we now develop takes as its starting point the failure of that of Mackintosh. Like its predecessor it accepts that the effectiveness of a CS will change according to its predictive power. It has two novel features as applied to excitatory conditioning. First, it suggests a new principle for determining changes in α; second, and more fundamentally, it abandons altogether the notion that changes occur in US effectiveness as conditioning proceeds. We begin by presenting our new account in a relatively informal way.

Adopting the perspective suggested by some models of information processing (Wagner, 1978) leads to the suggestion that associative learning depends on the conjoint processing of CS and US in some limited capacity device. A change in the associability of a stimulus is then represented in terms of the likelihood of its gaining access to this processor. We propose that stimuli such as the USs used in typical conditioning procedures are always likely to gain access to the processor. But for the processor to be used efficiently, access for other stimuli will be limited so that only those needed for learning gain entry. Stimuli that fully predict their consequences will be denied access to the processor, whereas stimuli that have recently been followed by surprising or unexpected events will receive processing. In other words (and in direct contrast to Mackintosh's model), we suggest that a stimulus is likely to be processed to the extent that it is not an accurate predictor of its consequences.

A More Formal Statement

The suggestion that the associability of a stimulus depends on how well it predicts a subsequent US may be expressed as follows:

$$\alpha_A{}^n = \left|\lambda^{n-1} - V_A{}^{n-1}\right|. \quad (8)$$

Here $\alpha_A{}^n$ represents the associability that a CS (A) will have on a given conditioning trial n. It depends on the absolute value of the discrepancy between the intensity of the US (λ) on the previous trial ($n - 1$) and the associative strength (V_A) of the stimulus on that trial. To conform to the convention that α for a stimulus varies between 0 and 1, we stipulate that λ may also vary only between these values.

Next, we wish to express formally the fact that the associative strength gained by a stimulus as a result of a conditioning trial depends not only on the associability of the CS but also on its intensity and that of the US. This is represented in Equation 9.

$$\Delta V_A = S_A \alpha_A \lambda, \quad (9)$$

where S is a parameter (varying between 0 and 1) that depends on CS intensity. The role played by the intensity of a stimulus is thus explicitly separated from that played by changes in its associability. Substituting for α in Equation 9 produces Equation 10.

$$\Delta V_A{}^n = S_A \left|\lambda^{n-1} - V_A{}^{n-1}\right| \lambda^n, \quad (10)$$

which gives the change in associative strength accruing to Stimulus A on conditioning trial n.

Equation 10 cannot be used to describe the changes in associative strength that occur on the first trial in which a CS is presented (and sizable changes can indeed occur on a single trial; see, e.g., Mahoney & Ayres, 1976). We are willing to leave this starting value for α unresolved for the time being, however, since it does not influence the predictions we wish to make. It should be added that Mackintosh's (1975b) model of changes in associability allows that such changes may generalize to some extent from one stimulus to another. If we accept this suggestion, then it follows that in any given experiment the starting value of α for a supposedly novel stimulus may well be set by the experience that the subjects have had with other similar stimuli.

Application to Simple Acquisition

All of the theories of conditioning discussed thus far have supposed that associative strength reaches an asymptote as the US becomes predicted and loses its effectiveness. The model just outlined rejects this idea; instead, the course of conditioning is determined solely by changes in CS effectiveness. On trials after the first, the associability of the CS will be determined by Equation 8, and thus, as conditioning proceeds, α will decline from a value close to that set by λ until, when the associative strength of the CS is equal to λ, α will be zero and there will be no further change in the strength of the association.

Although there is some variation from one experimental paradigm to another (Mackintosh, 1974, p. 9), the typical learning curve is usually taken to be sigmoid in shape. Relatively small increments in associative strength as the asymptote is approached are to be expected on the basis of the preceding equations. The occurrence of small increments at the start of conditioning might be taken to indicate that the starting value of α is low, but this need not necessarily be the case. The starting value of α could conceivably be high, but the failure to observe large increments in the magnitude of the conditioned response on the initial trials may be due to some threshold that associative strength must cross before it is translated into performance (cf. Mackintosh, 1974, p. 11).

Latent Inhibition and Related Phenomena

The model is applied to latent inhibition as follows: When a novel stimulus is presented in the absence of a reinforcer, λ will be 0, and, since the associative strength of the stimulus is also 0, the associability of the stimulus will decline. When the stimulus is subsequently paired with a US, no learning will occur on the first conditioning trial, and in this way a difference will be established between the associative strength acquired by a preexposed stimulus and that acquired by a novel stimulus.[4]

Our model also allows that latent inhibition of a sort will go on during conditioning itself. Recall the experiment by Hall and Pearce (1979) mentioned earlier, which showed that pretraining in which a CS signaled a weak shock reduced the readiness with which the CS subsequently formed an association with a stronger shock. Our interpretation of this finding is that α for the CS declines to zero during conditioning as the associative strength of the CS approaches the λ value determined by the intensity of the weak shock. It follows that at least on the first trial of training with the stronger shock, there will be no increase in the associative strength of the stimulus. Only after this trial, when the increase in shock intensity had produced a discrepancy between λ and V, will further learning occur.

This account of the phenomenon reported by Hall and Pearce generates a novel prediction. It suggests that learning occurs slowly during the second phase of conditioning with the stronger shock because initially the value of $|\lambda - V|$ is close to zero. Accordingly, some procedure that establishes a discrepancy between λ and V before the start of this second phase should allow further conditioning to proceed normally. For example, inserting one or a few trials on which the CS is not followed by shock between the phase of training when the CS predicts a small shock and the phase when it predicts a stronger shock should have this effect. Such a procedure will cause the value of V to exceed λ (which will be zero), and the associability of the CS will be increased. In other words, our model makes the seemingly paradoxical prediction that a phase of training that might be expected to teach the subject that the CS is not followed by shock could in fact help in forming the CS–strong shock association.

We have recently confirmed this prediction in a series of experiments using the conditioned suppression technique (Hall & Pearce, Note 2). In one of these experiments, two control groups received extensive initial training with a shock of moderate intensity as the US. The CS was a tone for

[4] It may be observed that this model predicts that the development of latent inhibition would be a much more rapid affair than is typically the case (Lubow, 1973). Discussion of this point will be deferred until exposition of the full model has been completed.

one group and a light for the other group. The two groups then received four test trials on which the tone signaled a stronger shock. It was found that the group that was pretrained with the light learned more rapidly during the test phase, thus replicating the results reported by Hall and Pearce (1979). A third group of subjects also experienced the tone CS during the pretraining and test phases, but, in addition, these animals received two nonreinforced presentations of the tone immediately before the test trials. According to the argument just outlined, these nonreinforced presentations of the CS should restore its associability and thus minimize the latent inhibition effect induced by pretraining with tone. We found that these experimental subjects acquired suppression during the test phase as readily as the control subjects that had received their pretraining with the light.

Conditioning With a Compound CS

We have thus far worked with the assumption that the associability of a CS is set by the discrepancy between its associative strength, V, and the strength of the US, λ. But we need now to specify how α changes as a result of trials when more than one CS is present, each of which may possess some associative strength. Given our central hypothesis that a stimulus will gain access to the processor only when it is needed for associative learning, that is, only when it has been followed by a surprising event, it seems sensible to assume that the associability of each stimulus in a compound will be determined by how well the reinforcer is predicted by the aggregate associative strength of all the stimuli in that compound. Provided the reinforcer is successfully predicted on some basis, there is no need for further learning about any CS. We suggest, therefore, that the value of a stimulus is given by the expression $|\lambda - V_\Sigma|$, where as before, V_Σ is the sum of the associative strengths of all stimuli present.

This model deals with blocking quite successfully. Given sufficient pretraining, the associative strength (V_B) of Stimulus B will rise to equal λ. Thus, in the normal blocking procedure when Stimulus A is added and the reinforcement parameters are not changed, the reinforcer will be fully predicted by B. The associability of the added element A for all but the first trial will be zero, that is, $|\lambda - V_\Sigma|$, since in this case V_Σ will equal V_B. On the first trial the associability of A will be at its normal starting value, and learning will occur on just this trial.

If the reinforcement parameters are changed to produce an increase in λ for the compound trials, conditioning to A will continue for a number of trials, since there will be a discrepancy between λ and V_Σ. The case of unblocking produced by the surprising omission of a second shock (Dickinson et al., 1976) needs a more elaborate explanation. In these circumstances the presence of the fully predicted first shock will produce a decline in α_A, whereas the omission of the second shock will produce an increase. As a result, α_A will not decline, or at least will not decline as much as it will for control subjects receiving both shocks. On subsequent trials this relatively high value of α_A will allow A to form an association with the first shock. It should be added that our account of inhibitory learning (to follow) allows that A may also form an association with the absence of the second shock, which may tend to attenuate the effects of the association with the first shock. We assume, however, that in this case the close temporal proximity of the first shock to A will ensure that excitatory learning will predominate over inhibitory learning, and this is enough for our present purposes. Further consideration of the relationship between shock omission, associability, and inhibitory learning will be given in the next major section.

The model can also be applied on the whole successfully to another much studied sort of compound conditioning, overshadowing, although here a new problem is encountered. Overshadowing is said to occur when the associative strength acquired by an element of a compound stimulus is less than that produced by conditioning this stimulus in isolation. According to our account, when two stimuli are conditioned as a compound, both will gain associative strength (perhaps at different rates, depending on their intensities) until their joint strengths equal λ. At

this point the associability of each will fall to zero and no further learning will occur. Neither stimulus will acquire as much strength as it would if conditioned in isolation. This prediction is supported by some of the experimental findings, but there are others (Mackintosh, 1976) which suggest that when the two stimuli differ markedly in intensity or salience, the more salient element overshadows the less salient, but not vice versa. If further research were to support this latter finding, then it would be necessary for us to modify our model. We could suggest, for instance, that the value of α on any given trial is determined by the value of $|\lambda - V_{max}|$ for the preceding trial, where V_{max} is the associative strength of the stimulus that most accurately predicts the US. This formulation retains our central assumption that the associability of a CS falls off when the US is accurately predicted, it leaves unchanged our analysis of blocking, and it predicts that the most salient element in an overshadowing experiment will reach full strength, since learning will stop only when this stimulus predicts the reinforcer. This formulation also predicts that overshadowing will not occur between stimuli of equal salience, a prediction for which there is, unfortunately, only limited support (Mackintosh, 1976). Accordingly, we intend to retain, for the time being, our initial assumption whereby α is set by the value of $|\lambda - V_{\Sigma}|$.

There is some evidence to show that overshadowing effects can occur when only a single trial of compound conditioning is given (Mackintosh, 1971; Mackintosh & Reese, 1979). Such an effect is predicted neither by our model nor by any of the other formal models that we have considered (including that of Mackintosh, 1975b). We must allow, therefore, that there is some process that can produce overshadowing different from that envisaged by our model. This process may also be active during experiments in which several compound training trials are given and may contribute to the overshadowing that is observed in these. If so, we can expect a perfect fit between data and theory only when this extra process is properly understood and incorporated into our theories.

A Conceptual Framework

So far we have presented our model only in a rather abstract way with the intention of showing that our equations fit the phenomena (and, indeed, predict new ones). It is appropriate at this stage, before turning to the rather different problems posed by inhibitory learning, to set our notions in the context of current, rather more general views of learning.

We began with a conceptual framework (essentially that of Konorski, 1948), which suggested that excitatory conditioning procedures result in the formation of an association between internal representations of the CS and the US such that presentation of the CS becomes capable of producing activity appropriate to the occurrence of the US. This account fits very well with those theories which suggest that the US becomes less effective as conditioning proceeds. Konorski, for instance, suggested that an association between CS "center" and US "center" was strengthened only when the CS was followed by an increase in activity in the US center caused by the occurrence of the US. Since the CS is itself capable of exciting the US center, it follows that once the CS has acquired sufficient associative strength, the level of activity aroused by the CS will be equivalent to that aroused by the US, and thus the reinforcer will be rendered ineffective. Wagner's (1976, 1978) model employs essentially the same mechanism. In his case, a CS will "prime" a representation of the US into a limited-capacity processor, making the processor less likely to accept information about the US when it actually occurs.

While accepting some features of these models (in particular, that an association is formed when effective representation of the CS and US both gain access to some processing device), the novel features of our theorizing prompt us to adopt a slightly different conceptual framework. We want to express the idea that an effective US representation is always present in the processor when the US occurs. In addition, we want to ensure that although the CS will continue to evoke the CR, the likelihood of an effective representation being present in the

processor will change according to how well the US has been predicted by that stimulus in the past. We require, therefore, a mechanism whereby a comparison can be made between the US actually presented (and thus represented in the processor) and that predicted by the associative strength of the stimulus. Such a mechanism is shown in Figure 1.

The figure depicts a processor containing representations of a CS and of a US on trial n of a simple conditioning procedure. The effectiveness or strength of the US representation is determined by the value of λ, that of the CS, by $S \cdot \alpha^n$. Since α^n will have been determined on trial $n - 1$, an "α store" is required to enable the processor to remember the value for trial n. The conjoint processing of the CS and US representations results not in the formation of a direct link between them but in an increase in the ability of the CS to excite what we may call a "US memory." The size of this increase will depend on the strengths with which CS and US are represented in the processor. The extent to which the US memory is excited on a given trial will depend on the associative strength of the CS (in fact, on V_Σ when more than one CS is present), and will be directly related to the strength or probability of the CR. Further, a comparison can be made, by means of the comparator, between the US memory and the representation of the US in the processor with the magnitude of any discrepancy being used to determine the associability of the CS on its next presentation. When there is no discrepancy and α falls to zero, the CS will no longer be represented in the processor; it will, however, remain capable of exciting activity in the US memory and thus of producing a CR.

This framework will be extended and discussed in more detail shortly; however, we turn now to an attempt to apply these general principles to the learning produced by inhibitory conditioning procedures.

Inhibitory Learning

The Problem

The problem we face in supplying an adequate account of inhibitory learning is rather more fundamental than that met when

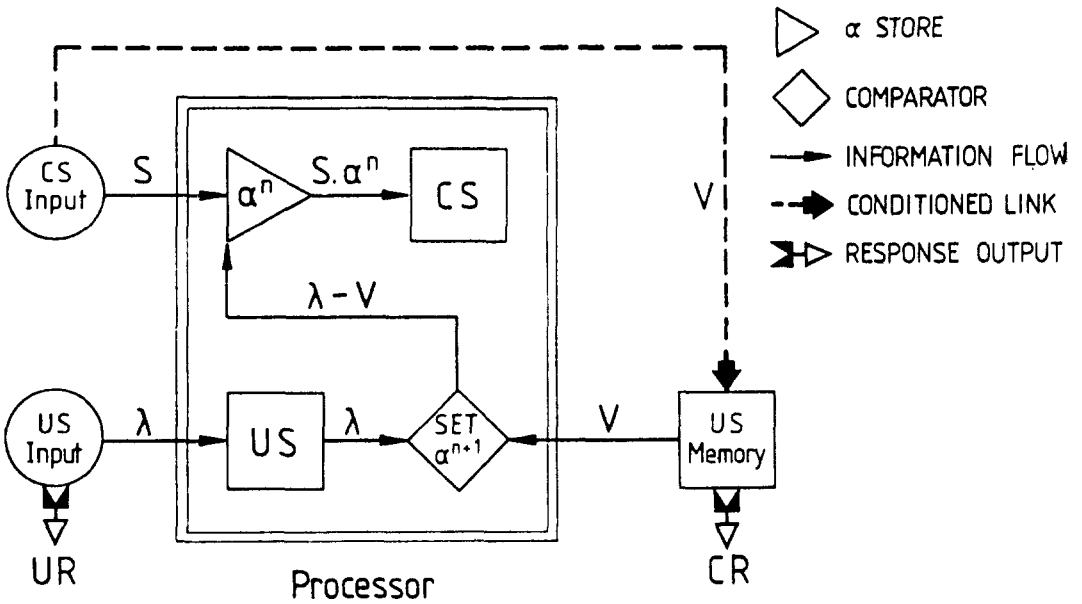

Figure 1. A possible information flow diagram for excitatory conditioning on trial n (a trial other than the first trial with the conditioned stimulus [CS]). (Symbols alongside links indicate the level of activity, where appropriate, induced in the representations to which they lead. US = unconditioned stimulus [see text for further explanation].)

we first considered excitatory learning. In that case there was, at least, fairly general agreement about the way in which the relationship between CS and US is represented internally. There is no such agreement in the case of inhibitory learning.

A development of the Rescorla-Wagner model (Wagner & Rescorla, 1972) exploited the way in which a negative correlation between CS and US can produce a CS with a negative V value to produce a highly successful account of the acquisition of conditioned inhibition and of the way in which inhibitors and excitors interact. But this model encounters a number of difficulties, not least the fact that it is difficult to conceptualize the nature of a negative V. Perhaps this problem could simply be shelved, but there are experimental data that run counter to the model's predictions, and these must be dealt with. For example, Zimmer-Hart and Rescorla (1974) have found that a stimulus established as an inhibitor maintains its inhibitory properties in spite of being presented repeatedly in isolation. According to the theory, such a CS has a negative V, and since it is presented in the absence of a US, λ will be zero. It follows from Equation 2 that there will be trial-to-trial increments in V until it, too, reaches zero and the associative strength and the asymptote are no longer discrepant.

Problems of this sort have prompted Rescorla (e.g., 1979) to accept an account of inhibition based on the work of Konorski (1948), an account that has the advantage of being open to a relatively simple psychological interpretation. The suggestion is that an inhibitory CS does not possess negative associative strength but rather functions by raising some threshold that an excitatory CS must exceed before its effects can become evident. This view not only solves the problem of why a nonreinforced inhibitor should fail to extinguish, but it also fits very well with the observation that inhibitors are largely ineffective in the absence of excitatory stimuli, hence the need for special tests to demonstrate their properties (Rescorla, 1969).

But we should not accept too readily the view that conditioned inhibitors lack response-eliciting powers. Consider the following experiment by Wasserman, Franklin, and Hearst (1974). (See also Hearst & Franklin, 1977). They established a lighted key as an inhibitor for pigeons by presenting food in its absence but not in its presence and demonstrated that the birds developed the behavior of withdrawing from the key when it lighted up. It is difficult to see how the raising of an excitatory threshold could produce an explicit withdrawal; instead we should give consideration to the idea that inhibitory training procedures may result in some independent form of learning that can, in certain circumstances, show itself directly in behavior.

A Possible Solution

Konorski (1967) has suggested that just as a positive correlation results in an association between representations of CS and US, so a negative correlation results in the formation of an association between CS and a representation of no US ($\overline{\text{US}}$). We intend to develop our own version of this view. It allows us to argue that inhibitory learning goes on when effective representations of a CS and of $\overline{\text{US}}$ are present in the processor, that is, to deal with inhibitory learning in much the same way as we dealt with excitatory learning. It is necessary, therefore, to say a little more about the nature of a $\overline{\text{US}}$ representation, about what events activate it, and how it influences behavior once activated.

We first assume that a $\overline{\text{US}}$ representation is activated only by the omission of an *expected* US, that is, one that is predicted by some excitatory CS. There are good precedents for the suggestion that the omission of an expected reinforcer is an effective event. Thus, Amsel (1958) suggested that the nonoccurrence of expected food will generate the emotional state of frustration; similarly, Denny (1971) holds that the nonoccurrence of anticipated shock will produce the emotional response of relaxation. These states may be regarded as unconditioned responses produced by a $\overline{\text{US}}$. The experiments of Wasserman et al. (1974) show that previously neutral stimuli associated with these states may become capable of evoking overt behavior. A stimulus

associated with the state of frustration produced by the omission of expected food can come to evoke a conditioned withdrawal response. It is possible that inhibitory stimuli for USs other than food may be capable of eliciting CRs, but the appropriate tests for demonstrating such a phenomenon have yet to be done.

We shall assume, therefore, that the occurrence of a $\overline{US}$ will, as with any excitatory US, gain automatic access to the processor and be processed along with any CS representation that is currently activated. The outcome of such processing will be an association between the CS and what we refer to as a $\overline{US}$ memory. The subsequent arousal of this memory by an inhibitory CS may lead to the occurrence of an inhibitory conditioned response ($\overline{CR}$).

A cardinal feature of inhibitory CSs is their ability to diminish the CRs elicited by excitatory stimuli for the same US, a feature that has led to the development of the retardation procedures (Rescorla, 1969) as tests for the presence of conditioned inhibition. To account for this property

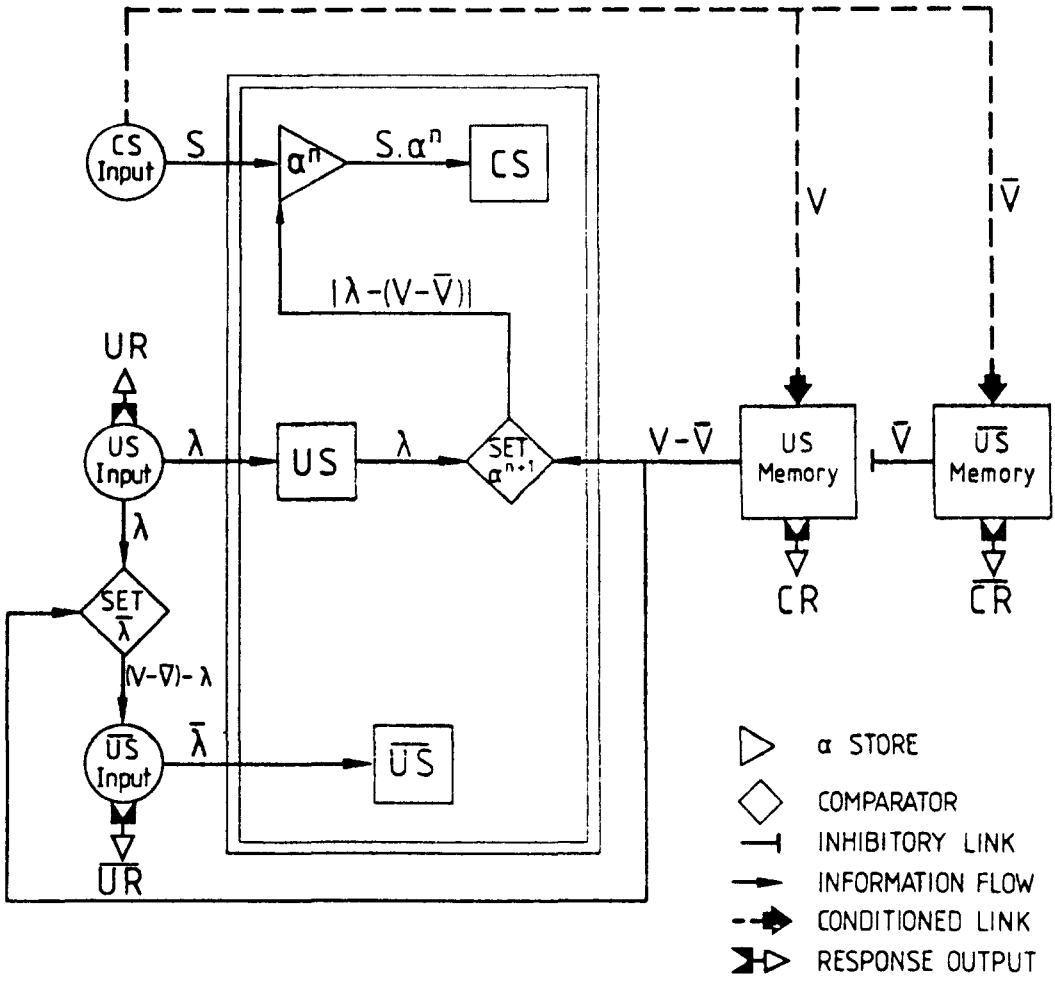

Figure 2. A possible information flow diagram for both excitatory and inhibitory conditioning on trial *n* (a trial other than the first trial with the conditioned stimulus [CS]). (Symbols alongside links indicate the level of activity, where appropriate, induced in the representations to which they lead. US = unconditioned stimulus [see text for further explanation].)

we again follow Konorski (1967) and propose that there exists an inhibitory relationship between US and the $\overline{US}$ memories for a given reinforcer, so that if they are both excited simultaneously, activity in the $\overline{US}$ memory will inhibit activity in the US memory. Since the level of activity in the US memory is directly related to the strength of the CR it elicits, it follows that presenting the appropriate inhibitory CS with an excitatory CS can reduce the strength of the response elicited by the latter. This relationship is expressed in the right-hand portion of Figure 2, an extended version of Figure 1, that also includes the mechanism, still to be discussed, for determining the magnitude of $\overline{US}$.

A Formalization

Our account of excitatory learning states that the increase in associative strength produced by a conditioning trial is determined by the salience of the CS, its associability, and the intensity of the reinforcer. We intend to apply these ideas almost unaltered to the inhibitory case. We do this in Equation 11:

$$\Delta \bar{V}_A = S_A \cdot \alpha_A \cdot \bar{\lambda}, \tag{11}$$

which is simply a rewritten version of Equation 9, the equation determining the course of excitatory conditioning. The value $\Delta \bar{V}_A$ represents the increment in the strength of a CS–$\overline{US}$ association for CS, A; $\bar{\lambda}$ represents the intensity of the reinforcer, the omission of the expected US. We have stated that the activation of a $\overline{US}$ representation will inhibit that of a US representation. In the absence of any direct evidence, we shall make the most straightforward assumption that a simple subtractive relationship governs this inhibitory interaction, that the value $\bar{V}$ should be subtracted from the value V, the strength of which determines the strength of any excitatory CR.

Intensity of the inhibitory reinforcer. The intensity of the excitatory reinforcer is given directly as λ; that of the inhibitory reinforcer requires more discussion, since $\bar{\lambda}$ is not directly produced by an environmental event but is rather the frustration (or relief) produced when some anticipated event fails to occur. Having said that the $\overline{US}$ representation is active when a predicted US fails to occur, we assume that the strength of this activation will be determined by the strength of the US that is predicted; that is, $\bar{\lambda}$ will depend on the associative strengths of such excitatory stimuli as are present (V_Σ). Experimental studies of a phenomenon that has been called "overexpectation" (Kremer, 1978; Rescorla, 1970; Wagner, 1971) allow us to specify this relationship more precisely. These studies have shown that when two stimuli that have each been separately paired with a US are presented in compound along with the same US, there is a reduction in the measured associative strength of each element. Moreover, the study by Kremer (1978) shows that a novel stimulus introduced on the compound trials requires inhibitory properties. It appears, therefore, that the effective reinforcer for inhibitory learning is not simply the absence of the US but rather the absence of a US as intense as that predicted. Given the inhibitory relationship between V and $\bar{V}$, the predicted intensity of the US will be given by $(V_\Sigma - \bar{V}_\Sigma)$ when stimuli associated with the $\overline{US}$ memory are present. The value of $\bar{\lambda}$ seems therefore to be set according to an equation of the sort given below:

$$\bar{\lambda} = (V_\Sigma - \bar{V}_\Sigma) - \lambda. \tag{12}$$

Obviously, in normal inhibitory conditioning procedures, where the US is not presented, $\bar{\lambda}$ will simply be equal to $V_\Sigma - \bar{V}_\Sigma$, and given this we shall still refer to no-US representations.

Returning to Figure 2, it is evident from Equation 12 that the magnitude of the $\overline{US}$ input is given by the discrepancy between the activity in the US memory and that in the US input. This has been represented by placing a comparator between these two centers, the output of which feeds into the $\overline{US}$ input. We must obviously stipulate that this comparator will be operative only when activity in the US memory exceeds that in the US input.

Associability changes in inhibitory conditioning. In our earlier discussion of the

experiment by Hall and Pearce (Note 2) we have already pointed out how the omission of a predicted shock can restore the associability of a stimulus. This follows directly from Equation 8, which specifies that α is set by the discrepancy between the intensity of the US and the intensity predicted by the CS. The present discussion enables us to refine this formulation. We have argued that a discrepancy between the intensity of the US and the predicted intensity will establish a representation of $\overline{\text{US}}$ (but only when V_Σ exceeds λ) and thus allow CS – $\overline{\text{US}}$ processing to occur. Accordingly, the strength of any such association must be taken into account in determining the value of α, and we may rewrite Equation 8 in the more general form of Equation 13:

$$\alpha_A{}^n = \left|\lambda^{n-1} - (V_\Sigma{}^{n-1} - \bar{V}_\Sigma{}^{n-1})\right| \quad (13)$$

Note that the value of the expression given in Equation 13 will be numerically equivalent to the value for the inhibitory reinforcer, $\bar{\lambda}$, as given by Equation 12. That is to say that the size of the inhibitory reinforcer depends on the amount of relief produced by the omission of an expected shock, for example; likewise, the associability of a stimulus for the next trial, which we have said will depend on the difference between what is predicted on a given trial and what actually happens, is also determined by the amount of relief that occurs.

Application of the Model

Extinction

The proposed view of inhibition leads to an interpretation of extinction that differs from that adopted by certain other accounts of conditioning. These (e.g., Rescorla & Wagner, 1972) suggest that extinction consists of the weakening of previously established associations, whereas we regard extinction as a new form of conditioning, the reinforcer being $\overline{\text{US}}$. The existing CS–US association is not lost as a result of the omission of the US. Rather, a new CS–$\overline{\text{US}}$ association is formed, which by virtue of the inhibitory link between the US and $\overline{\text{US}}$ memories leads to a reduction in the strength of the CR. Learning during extinction, according to this model, will cease for two reasons. As the strength of the inhibitory association, $\bar{V}$, increases with continued nonreinforcement, the difference between V and $\bar{V}$ will approach zero, at which point α, according to Equation 13, and $\bar{\lambda}$, according to Equation 12, will also equal zero. Learning will thus cease both because the inhibitory reinforcer is no longer present and because the CS is no longer permitted access to the processor.

This model makes an interesting prediction about the course of learning during the initial trials of extinction. If during previous excitatory conditioning the associative strength of the CS has come near to asymptote, then its associability will be low on the first extinction trial, and little inhibitory learning can be expected. It should be possible, however, to enhance the rate of extinction by making some change in the reinforcement parameters on the last trial of acquisition, since such a change would restore the associability of the CS and would allow rapid learning to take place on the immediately succeeding trials. Indeed, even an increase in the size of the reinforcer should hasten the course of subsequent extinction (just as, in the experiment by Hall & Pearce, Note 2, omission of an expected shock was shown to be capable of enhancing further excitatory learning).

As a test of this prediction, we have carried out an experiment (Hall & Pearce, Note 3) in which two groups of rats were given conditioned suppression training in which a tone was paired with a relatively weak electric shock. One (control) group then received extinction trials with the tone presented alone. The second ("surprise") group received an unexpected stronger shock on its last conditioning trial. We predicted that extinction would proceed rapidly for the second group, since the discrepancy between V and λ established on the last trial of acquisition will restore the lost associability of the CS.

The results confirmed this prediction. On the first extinction trial, responding was slightly more suppressed by the tone in the surprise group than in the control group. As extinction progressed, however, there was a

more rapid loss of suppression by the surprise group so that these subjects were significantly less suppressed than the control group on the later trials of extinction.

Conditioned Inhibition

In the procedure typically used to establish conditioned inhibition, one stimulus (A) is established as an excitor and then a second stimulus (B) is added, the AB compound not being followed by the reinforcer. The acquisition of inhibition by B is explained in much the same way as in the acquisition of inhibition during the simple extinction of an excitatory CS. It differs only in that the associability of the stimulus in question will in this case already be quite high at the beginning of inhibitory training, also, in that the activity in the US memory (V), which is responsible for the activation of $\overline{\text{US}}$, is brought about by the presence of the other stimulus, A.

We should also consider the case in which a third stimulus, C, that has already been established as an inhibitory CS, is added to the AB compound, the whole being presented without being followed by a US. In this case the omission of the US will be predicted by Stimulus C. Indeed, if the associative strength of Stimulus A for the US and the associative strength of Stimulus C for $\overline{\text{US}}$ are roughly the same, it follows from Equation 13 that the associabilities of the stimuli present will fall to near zero, as a result of the first nonreinforced compound trial. This will be true for Stimulus B, and thus little strength will be acquired by B as a signal for $\overline{\text{US}}$ on subsequent compound trials. In other words, the presence of a stimulus that predicts nonreinforcement should be able to block the acquisition of inhibition by Stimulus B. This effect has been demonstrated experimentally by Suiter and LoLordo (1971).

Our account meets no difficulty with the fact that presentation of an inhibitor in isolation does not extinguish its inhibitory power. The view being proposed here is that extinction occurs when a CS enters into an association with a representation antagonistic to that underlying acquisition. Thus, the extinction of a conditioned inhibitor will occur only in circumstances that ensure the formation of an association with some representation of the US.

Supernormal Conditioning

By virtue of the fact that activation of the $\overline{\text{US}}$ memory influences the activity of the US memory, it is possible for the presence of an inhibitory CS to influence the course of excitatory conditioning. In particular, excitation of the $\overline{\text{US}}$ memory will suppress activity in the US memory, and as a result, any discrepancy between the US memory and the US actually presented will be greater than normal. Since this discrepancy determines the associability of the excitatory CS on subsequent trials, it follows that the development of the associative bond will progress more rapidly and reach a higher asymptote when conditioning takes place in the presence of an inhibitory CS than it would in the absence of such a CS. It is gratifying to note that this type of result has been found in experiments on "superconditioning" (Rescorla, 1971).

Summary and General Comments

In the preceding sections we have developed and modified our original model. It is appropriate at this point, therefore, to give a concise and fairly abstract summary of the model in its final form.

The model is essentially one that describes how the likelihood of a CS (A) being processed (α_A) changes as a result of experience. The value of α_A on a given trial n is given by the absolute value of the discrepancy between the intensity of the US occurring on the previous trial and the extent to which this US is predicted on that trial. Thus,

$$\alpha_A{}^n = \left|\lambda^{n-1} - V_T{}^{n-1}\right|, \qquad (14)$$

where λ represents the intensity of the US, and V_T the aggregate associative strength of all stimuli present on that trial; that is, $V_\Sigma - \bar{V}_\Sigma$, where V_Σ is the total associative strength of all excitatory CSs and $\bar{V}_\Sigma$ is

the total associative strength of all inhibitory CSs.

When the CS has an α value that assures that it receives processing, the change in associative strength that it undergoes on a given trial is determined according to Equation 9, where S_A is the salience of the CS, λ the intensity of the US, and α_A is given by events on the previous trial as in Equation 13.

When a US is omitted or is less intense than that predicted by the stimuli that are present, the inhibitory reinforcer ($\bar{\lambda}$) is held to occur, its value being given by Equation 12:

$$\bar{\lambda} = (V_\Sigma - \bar{V}_\Sigma) - \lambda.$$

Inhibitory learning then occurs according to Equation 11.

Further Factors Governing CS Associability

According to the model as it has been presented, the associability of a CS is determined exclusively by the events that occurred on the immediately preceding trial. This simplification allowed us to present the model in a relatively straightforward way, but it is now necessary to discuss other factors that are likely to determine the associability of a stimulus.

First, it seems likely that events on several, perhaps many, of the preceding trials will influence the associability of a CS on a given conditioning trial. This would serve the purpose of minimizing the disruptive influence of chance or atypical CS–US pairings. We might assume, for instance, that the "α store" shown in the figures stores not just the value specified by the immediately preceding trial but also the values resulting from some fixed number (c) of earlier trials. The value of α on trial n would then be given by some average of these values as in Equation 15,

$$\alpha^n = \frac{1}{c} \sum_{n-c}^{n-1} |\lambda - V_T|, \qquad (15)$$

or perhaps by some more complex averaging system that gave greater weight to more recent than to more remote trials.

Extending the model in this way does not radically alter its application to the experiments considered earlier. Thus, the acquisition of excitation or of inhibition will take place as originally specified, but Equation 16 means that the changes in associability will be less rapid. The analysis of blocking also will not be affected. Only in the case of latent inhibition does the modification embodied in Equation 15 markedly change the predictions of the model. In discussing this phenomenon we noted that the model incorrectly predicts that one nonreinforced presentation of a neutral stimulus is sufficient to produce maximal latent inhibition. In contrast, by using Equation 15, the model predicts that the effects of latent inhibition will be determined by a number of reinforced presentations of the CS. Since α^n reflects a mean value computed from $|\lambda - V_T|$ for each of the preceding trials, it follows that the greater the number of such trials on which $|\lambda - V_T|$ is equal to zero, the more slowly will α increase during subsequent conditioning.

Next, we must acknowledge the possibility that factors other than its immediate history of conditioning may determine the associability of a CS. There is increasing evidence that the nature of the US may influence CS associability. The bulk of this evidence comes from work on conditioned taste aversions where it has been shown that USs that produce illness may be preferentially associated with some CSs rather than others, but there is also evidence (e.g., see LoLordo, 1979) that effects of this sort are not restricted to the taste aversion paradigm. These effects may not be the result of the subjects' past experience with the stimuli (but see Mackintosh, 1973), and if so, they cannot be expressed within our model by changes in the parameter α. Perhaps, then, the parameter S, in addition to being determined by the intensity of the CS, is also influenced by the nature of the US. S may be low for an auditory CS for pigeons when the US is food but high when the US is electric shock. Of course, this is by no means a full explanation, but it does make clear that our model allows a distinction between two kinds of associability. On the one hand there are conditioned changes in associability (in α) brought

about by surprising or expected events; on the other, there are differences in associability that are more enduring in nature and are evident from an early stage in the animal's development. It may be necessary to invoke this second sort of associability to explain the recent findings of Testa (1975) and of Rescorla (e.g., Rescorla & Cunningham, 1979; Rescorla & Furrow, 1977) that learning may be facilitated when the events to be learned about are closely similar or are spatially contiguous.

Associability, Learning, and Performance

We do not intend to discuss the way in which activation of a US (or $\overline{US}$) representation produces changes in behavior; rather, we want to elucidate the relation between the associability of a CS and the performance of the CR.

We have taken the associability of a CS as being a measure of the likelihood that it will gain access to some limited capacity processor. Ideas of this general sort have been common in models of human information processing (e.g., Broadbent, 1958) and have been transferred to recent models of associative learning in animals (e.g., Wagner, 1976, 1978). The animal models have uniformly assumed, however, that a stimulus must gain access to the processor if it is to elicit any response. But this is an assumption that we cannot make for a number of reasons. To take the simplest, our model suggests that when the associative strength of a CS has reached asymptote so that the probability or strength of the CR is at its maximum, the associability of the CS will be zero. Thus, we suggest that a stimulus may evoke a response even though it fails to engage the mechanisms concerned with associative learning.

This suggestion receives some support from more recent accounts of human information processing (LaBerge, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). These theorists have adopted the view that alternative strategies may be available to the subject when a stimulus is presented. A *controlled* processing strategy will be used when the stimulus is novel or is presented in the context of a relatively unfamiliar task. This strategy permits the subject to learn about the stimulus and its relationship to other events, and it is processing of this sort that requires the use of the limited capacity processor. But once the subject is familiar with the task, it is assumed that an automatic processing strategy can be used that bypasses the central processor. In this terminology, the loss of associability of a CS is regarded as a transition from controlled to automatic processing.

Evaluation of The Model

In the first part of this article, we reviewed a number of recent theories of learning and outlined some of the evidence that we considered to be incompatible with them. Our model possesses the advantage of being able to overcome these difficulties, but how does it fare in a more general comparison with these theories? On a favorable note, the model was able to generate two novel predictions that we have confirmed, and it is not clear that these results are compatible with other theories concerned with changes in associability. In both experiments we demonstrated that a surprising event can restore the associability of a CS. This is clearly not consistent with Mackintosh's (1975b) theorizing, since the CS should not have lost associability in the first place. Moreover, recall that Wagner (1976, 1978) maintained that the associability of a CS is inversely related to the strength of an association between the CS and the context. It is very difficult to see how a surprising shock omission (Hall & Pearce, Note 2) or shock increase (Hall & Pearce, Note 3) after a CS can reduce the strength of the association between the CS and the context and thus restore associability.

Having said this, we must admit that there are features of both Wagner's and Mackintosh's theories that enable them to explain data that are incompatible with our theory as it currently stands. Mackintosh (1973, 1975b) has argued that changes in associability are reinforcer specific, so that a CS may be high in associability for food but low in associability for shock. In our model this dual associability is not possible, and the model is therefore incapable of explaining

the recent finding by Dickinson and Mackintosh (1979). They demonstrated that the associability of a stimulus followed by surprising food remained high for future conditioning involving food but not for future learning involving shock. One way for our model to accommodate this result is to propose that there are separate processors for learning about different reinforcers such as food and shock.

A further problem is posed by the evidence showing that the low associability of a stimulus can be restored by presenting it in a novel context (e.g., Dexter & Merrill, 1969; Lantz, 1973; Lubow, Rifkin, & Alek, 1976). This result accords exactly with Wagner's theorizing but is inconsistent with our own, since we argue that associability can only be restored by pairing the stimulus with a surprising event. One method for explaining this finding can, however, be developed by assuming that the context in which a stimulus occurs forms a part of the information by which the stimulus is encoded. Thus, presenting a familiar stimulus in a novel context will be equivalent to presenting subjects with a different or novel stimulus, and thus we would expect its associability to be high.

We have proposed that the amount of learning is determined by the amount of simultaneous processing that the representations of the CS and US receive in the processor. We should not expect, therefore, that the occurrence of a surprising event soon after a conditioning trial would influence learning on that trial. But this is the sort of result that has been reported by Wagner, Rudy, and Whitlow (1973) and Kremer (1979). Wagner et al. suggested that their result occurred because the processor is of limited capacity and that the surprising events restricted the amount of rehearsal allotted to the preceding learning experience. If we are to accommodate this type of result, then it may be necessary to assume first that the processor in our model is also of limited capacity, and second, that learning about two events continues even after the termination of these events. However, the conditions leading to the overloading of the processor need to be specified precisely before we can resort with any justification to the sort of explanation provided by Wagner et al.

Reference Notes

1. Bolles, R. C., & Sigmundi, R. A. *CS familiarity reduces its conditionability*. Paper presented at the meeting of the Psychonomic Society, Phoenix, November 1979.
2. Hall, G., & Pearce, J. M. *Restoring the associability of a preexposed CS by a surprising event*. Manuscript submitted for publication, 1980.
3. Hall, G., & Pearce, J. M. *Restoring the associability of a stimulus by a surprising event*. Paper presented at the meeting of the Experimental Psychology Society, York, England, April 1979.

References

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, *55*, 102–119.

Bindra, D. A unified account of classical conditioning and operant training. In A. M. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory*. New York: Appleton-Century-Crofts, 1972.

Broadbent, D. E. *Perception and communication*. Oxford, England: Pergamon Press, 1958.

Denny, M. R. Relaxation theory and experiments. In F. R. Brush (Ed.), *Aversive conditioning and learning*. New York: Academic Press, 1971.

Dexter, W. R., & Merrill, H. K. Role of contextual discrimination in fear conditioning. *Journal of Comparative and Physiological Psychology*, 1969, *69*, 677–681.

Dickinson, A., Hall, G., & Mackintosh, N. J. Surprise and the attenuation of blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, 1976, *2*, 213–222.

Dickinson, A., & Mackintosh, N. J. Classical conditioning in animals. *Annual Review of Psychology*, 1978, *29*, 587–612.

Dickinson, A., & Mackintosh, N. J. Reinforcer specificity in the enhancement of conditioning by posttrial surprise. *Journal of Experimental Psychology: Animal Behavior Processes*, 1979, *5*, 162–177.

Frey, P. W., & Sears, R. J. Model of conditioning incorporating the Rescorla-Wagner associative axiom, a dynamic attention process, and a catastrophe rule. *Psychological Review*, 1978, *85*, 321–340.

Hall, G., & Pearce, J. M. Latent inhibition of a CS during CS–US pairings. *Journal of Experimental Psychology: Animal Behavior Processes*, 1979, *5*, 31–42.

Hearst, E., & Franklin, S. R. Positive and negative relations between a signal and food: Approach-withdrawal behavior to the signal. *Journal of Experimental Psychology: Animal Behavior Processes*, 1977, *3*, 37–52.

Kamin, L. J. Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior*. New York: Appleton-Century-Crofts, 1969.

Konorski, J. *Conditioned reflexes and neuron organization*. New York: Cambridge University Press, 1948.

Konorski, J. *Integrative activity of the brain*. Chicago, Ill.: University of Chicago Press, 1967.

Kremer, E. F. The Rescorla-Wagner model: Losses of associative strength in compound conditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 1978, *4*, 22–36.

Kremer, E. F. Effect of posttrial episodes on conditioning in compound conditioned stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, 1979, *5*, 130–141.

LaBerge, D. Acquisition of automatic processing in perceptual and associative learning. In P. M. A. Rabbit & S. Dornic (Eds.), *Attention and performance V*. New York: Academic Press, 1975.

Lantz, A. E. Effects of number of trials, interstimulus interval and dishabituation during CS habituation on subsequent conditioning in a CER paradigm. *Animal Learning and Behavior*, 1973, *1*, 273–277.

LoLordo, V. M. Selective associations. In A. Dickinson & R. A. Boakes (Eds.), *Mechanisms of learning and motivation*. Hillsdale, N.J.: Erlbaum, 1979.

Lubow, R. E. Latent inhibition. *Psychological Bulletin*. 1973, *79*, 398–407.

Lubow, R. E., Rifkin, B., & Alek, M. The context effect: The relationship between stimulus preexposure and environmental preexposure determines subsequent learning. *Journal of Experimental Psychology: Animal Behavior Processes*, 1976, *2*, 38–47.

Lubow, R. E., Schnur, P., & Rifkin, B. Latent inhibition and conditioned attention theory. *Journal of Experimental Psychology: Animal Behavior Processes*, 1976, *2*, 163–174.

Mackintosh, N. J. An analysis of overshadowing and blocking. *Quarterly Journal of Experimental Psychology*, 1971, *23*, 118–125.

Mackintosh, N. J. Stimulus selection: Learning to ignore stimuli that predict no change in reinforcement. In R. A. Hinde & J. Stevenson-Hinde (Eds.), *Constraints on learning*. London: Academic Press, 1973.

Mackintosh, N. J. *The Psychology of animal learning*. London: Academic Press, 1974.

Mackintosh, N. J. Blocking of conditioned suppression: Role of the first compound trial. *Journal of Experimental Psychology: Animal Behavior Processes*, 1975, *2*, 335–345.(a)

Mackintosh, N. J. A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, 1975, *82*, 276–298.(b)

Mackintosh, N. J. Overshadowing and stimulus intensity. *Animal Learning and Behavior*, 1976, *4*, 186–192.

Mackintosh, N. J., Bygrave, D. J., & Picton, B. M. B. Locus of the effect of a surprising reinforcer in the attenuation of blocking. *Quarterly Journal of Experimental Psychology*, 1977, *29*, 327–336.

Mackintosh, N. J., & Reese, B. One-trial overshadowing. *Quarterly Journal of Experimental Psychology*, 1979, *31*, 519–526.

Mahoney, W. J., & Ayres, J. J. B. One-trial simultaneous and backward fear conditioning as reflected in conditioned suppression of licking in rats. *Animal Learning and Behavior*, 1976, *4*, 357–362.

Pearce, J. M., & Hall, G. Loss of associability by a compound stimulus comprising excitatory and inhibitory elements. *Journal of Experimental Psychology: Animal Behavior Processes*, 1979, *5*, 19–30.

Rescorla, R. A. Pavlovian conditioned inhibition. *Psychological Bulletin*, 1969, *72*, 77–94.

Rescorla, R. A. Reduction in the effectiveness of reinforcement after prior excitatory conditioning. *Learning and Motivation*, 1970, *1*, 372–381.

Rescorla, R. A. Variations in the effectiveness of reinforcement and nonreinforcement following prior inhibitory conditioning. *Learning and Motivation*, 1971, *2*, 113–123.

Rescorla, R. A. Conditioned inhibition and extinction. In A. Dickinson & R. A. Boakes (Eds.), *Mechanisms of learning and motivation*, Hillsdale, N.J.: Erlbaum, 1979.

Rescorla, R. A., & Cunningham, C. L. Spatial contiguity facilitates Pavlovian second-order conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 1979, *5*, 152–161.

Rescorla, R. A., & Furrow, D. R. Stimulus similarity as a determinant of Pavlovian conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, 1977, *3*, 203–215.

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

Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, 1977, *84*, 1–66.

Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 1977, *84*, 127–190.

Suiter, R. D., & LoLordo, V. M. Blocking of inhibitory Pavlovian conditioning in the conditioned emotional response procedure. *Journal of Comparative and Physiological Psychology*, 1971, *76*, 137–144.

Testa, T. J. Effects of similarity of location and temporal intensity pattern of conditioned and unconditioned stimuli on the acquisition of conditioned suppression in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, 1975, *1*, 114–121.

Wagner, A. R. Elementary Associations. In H. H. Kendler & J. T. Spence (Eds.), *Essays in neobehaviorism: A memorial volume to Kenneth W. Spence*. New York: Appleton-Century-Crofts, 1971.

Wagner, A. R. Priming in STM: An information processing mechanism for self-generated or retrieval-generated depression in performance. In T. J. Tighe & R. N. Leaton (Eds.), *Habituation: Perspectives from child development, animal behavior, and neurophysiology*. Hillsdale, N.J.: Erlbaum, 1976.

Wagner, A. R. Expectancies and the priming of STM. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior*. Hillsdale N.J.: Erlbaum, 1978.

Wagner, A. R., & Rescorla, R. A. Inhibition in Pavlovian conditioning: Application of a theory. In M. S. Halliday & R. A. Boakes (Eds.), *Inhibition and learning*. London: Academic Press, 1972.

Wagner, A. R., Rudy, J. W., & Whitlow, J. W. Rehearsal in animal conditioning. *Journal of Experimental Psychology*, 1973, *97*, 407–426.

Wasserman, E. A., Franklin, S. R., & Hearst, E. Pavlovian appetitive contingencies and approach versus withdrawal to conditioned stimuli in pigeons. *Journal of Comparative and Physiological Psychology*, 1974, *86*, 616–627.

Zimmer-Hart, C. L., & Rescorla, R. A. Extinction of Pavlovian conditioned inhibition. *Journal of Comparative and Physiological Psychology*, 1974, *86*, 837–845.

Received October 15, 1979 ■

U S POSTAL SERVICE

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

(Required by 39 U.S.C. 3685)

1 TITLE OF PUBLICATION	2 PUBLICATION NO	3 DATE OF FILING
PSYCHOLOGICAL REVIEW	4 4 8 8 0 0	9/30/80
3 FREQUENCY OF ISSUE	4 NO OF ISSUES PUBLISHED ANNUALLY	5 ANNUAL SUBSCRIPTION PRICE
Bimonthly	6	$8 indiv./24 nonindiv.

6 LOCATION OF KNOWN OFFICE OF PUBLICATION *(Street, City, County, State and ZIP Code) (Not printers)*

1400 North Uhle Street, Arlington, VA 22201

7 LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS *(Not printers)*

1200 17th Street N.W., Washington, D.C. 20036

8 NAMES AND COMPLETE ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR

PUBLISHER *(Name and Address)*

American Psychological Association, 1200 17th St., N.W., Washington, DC 20036

EDITOR *(Name and Address)*

William K. Estes, Harvard Univ., 620 William James Hall, Cambridge, MA 02138

MANAGING EDITOR *(Name and Address)*

Anita DeVivo, 1400 North Uhle Street, Arlington, VA 22201

9 OWNER *(If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given. If the publication is published by a nonprofit organization, its name and address must be stated.)*

NAME	ADDRESS
American Psychological Association	1200 17th St., N.W., Washington, D.C. 20036

10 KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES *(If there are none, so state)*

NAME	ADDRESS
None	

11 FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES *(Section 132.122, PSM)*

The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes *(Check one)*

[X] HAVE NOT CHANGED DURING PRECEDING 12 MONTHS [ ] HAVE CHANGED DURING PRECEDING 12 MONTHS *(If changed, publisher must submit explanation of change with this statement.)*

12 EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A TOTAL NO. COPIES PRINTED *(Net Press Run)*	9,364	9,293
B PAID CIRCULATION 1 SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	- 0 -	- 0 -
2 MAIL SUBSCRIPTIONS	7,584	7,486
C TOTAL PAID CIRCULATION *(Sum of 10B1 and 10B2)*	7,584	7,486
D FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	178	170
E TOTAL DISTRIBUTION *(Sum of C and D)*	7,762	7,656
F COPIES NOT DISTRIBUTED 1 OFFICE USE, LEFT OVER, UNACCOUNTED, SPOILED AFTER PRINTING	1,602	1,637
2 RETURNS FROM NEWS AGENTS	- 0 -	- 0 -
G TOTAL *(Sum of E, F1 and 2—should equal net press run shown in A)*	9,364	9,293

13 I certify that the statements made by me above are correct and complete

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER

Anita DeVivo

13. FOR COMPLETION BY PUBLISHERS MAILING AT THE REGULAR RATES *(Section 132.121, Postal Service Manual)*

39 U.S.C. 3626 provides in pertinent part: "No person who would have been entitled to mail matter under former section 4359 of this title shall mail such matter at the rates provided under this subsection unless he files annually with the Postal Service a written request for permission to mail matter at such rates."

In accordance with the provisions of this statute, I hereby request permission to mail the publication named in Item 1 at the phased postage rates presently authorized by 39 U.S.C. 3626.

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER

PS Form 3526 Aug 1978 *(Page 1)* (See instructions on reverse)