Animal Learning & Behavior
2000, 28 (3), 211-246

An elemental model of associative learning: I. Latent inhibition and perceptual learning

I. P. L. McLAREN and N. J. MACKINTOSH
University of Cambridge, Cambridge, England

This paper presents a brief, informal outline followed by a formal statement of an elemental associative learning model first described by McLaren, Kaye, and Mackintosh (1989). The model assumes representation of stimuli by sets of elements (i.e., microfeatures) and a set of associative algorithms that incorporate the following: real-time simulation of learning; an error-correcting learning rule; weight decay that distinguishes between transient and permanent associations; and modulation of associative learning that gives high salience to and, hence, promotes rapid learning with novel, unpredicted stimuli and reduces the salience for a stimulus as its error term declines. The model is applied in outline fashion to some of the basic phenomena of simple conditioning and, in greater detail, to the phenomena of latent inhibition and perceptual learning. A detailed account of generalization and discrimination will be provided in a later paper.

In this paper, we attempt to show what associative, elemental models have to offer the learning theorist by considering one particular instantiation of this type of model, which was originally outlined some years ago (McLaren, Kaye, & Mackintosh, 1989). We begin with a brief introduction to the model, emphasizing the representational assumptions contained in it and giving an overview of the novel mechanisms it contains for salience modulation and trace decay. This introduction is followed by a more formal presentation of the model, and we then move on to consider the experimental evidence addressed by the model and, in particular, how it copes with data that have come to light in the last decade. In this paper, we focus mainly on latent inhibition and perceptual learning, leaving to a later paper a more detailed discussion of discrimination and generalization. Our conclusion is that, in general, the model does sufficiently well to establish just how much can be done with a simple, associative analysis, but there are areas where the model is either inadequate and requires further development or may well be just plain wrong. In any case, we believe that the representational assumptions and associative mechanisms considered here are likely to have some significant role to play in any future theory of associative learning.

AN OUTLINE OF THE THEORY

Representation of Stimuli

Conditioning theorists have not often taken much trouble to specify how the stimuli they use in their experiments might be represented. Since those stimuli are typically lights, tones, buzzers, flavors, and so forth, the question has not seemed to require any very complicated answer. The most common, perhaps unreflective, suggestion has been to take what we shall term an elemental stance, which sees stimuli as sets of elements and generalization between one stimulus and another coming about as a consequence of their sharing common elements. Early statements of this position were provided by Thorndike (1911) and Hull (1943), the idea was developed and formalized in statistical learning theory (Atkinson & Estes, 1963; Bush & Mosteller, 1951), and is incorporated without much comment into such standard modern models as Rescorla and Wagner (1972) and Wagner (1981). It was left to those who were more specifically interested in discrimination learning and choice to consider other possibilities. Spence (1952) allowed that although animals would normally represent discriminative stimuli in an elemental fashion, they were capable, if necessary, of representing them as compounds (a black door on the left or a white door on the right)—a point of view developed by Medin (1975). Gulliksen and Wolfle (1938) developed a fully fledged configural analysis of discrimination learning, according to which animals trained on a simultaneous black–white discrimination learned to make one response to the configuration of black on the left and white on the right and another to the configuration of black on the right and white on the left. Pearce (1987, 1994) has recently done much to revive interest in a configural approach to conditioning and discrimination learning.

The most explicit of the elemental theories of conditioning was stimulus-sampling theory (Atkinson & Estes, 1963; Estes, 1959), which conceptualized all stimuli as sets of elements, the elements themselves being simple primitives (corresponding, perhaps, to what would today be termed the microfeatures of a stimulus). The central postulate of the theory was that only a subset of the ele-

Correspondence concerning this article should be addressed to I. P. L. McLaren, Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, England (e-mail: iplm2@cus.cam.ac.uk).

—Editor's Note: This article was invited by the editors.

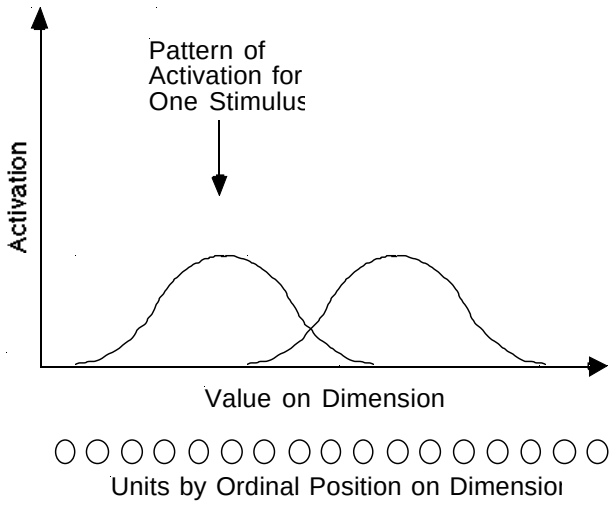

Figure 1. Representation of a stimulus dimension in an elemental system. A given point on the dimension is coded by a pattern of activation over units tuned to different points on that dimension. The pattern of activation indicated is taken to be approximately Gaussian and centered on the stimulus value on the dimension. See the text for more details.

ments potentially activated by the presentation of a given stimulus are actually sampled, and hence active, on any one occasion on which the stimulus is presented. This sampling postulate allowed the theory to combine the principle of all-or-none, one-trial learning with the observed reality that conditioning typically proceeds rather slowly. Although the elements actually sampled on one trial will all be fully conditioned on that trial, since they constitute only a random subset of the total, there will be significant variability in responding from one trial to the next, and conditioning will not be complete until virtually all the elements have been sampled and thus conditioned.

We take stimulus-sampling theory as our point of departure, but neither one-trial learning nor random sampling of elements are assumed in our theorizing. And we also part company with early versions of stimulus-sampling theory that, in true stimulus–response tradition, allowed associations to be formed only between stimuli and responses. The critical assumptions we borrow are the following.

1. The representation of a stimulus consists of a pattern of graded activation distributed over a set of units corresponding to the elements of the stimulus, rather than there being a one-to-one correspondence between a stimulus and a representational unit.

2. Similar stimuli consist of partially overlapping sets of elements, their degree of similarity being related to the proportion of common elements. Where stimuli can be construed as varying along a continuum or dimension, such as wavelength or auditory frequency, different values along the dimension are assumed to consist of a series of overlapping sets of elements (see Hull, 1943; Thompson, 1965). In effect, each representational unit is postulated to have a *tuning curve*, responding most strongly to one particular value on the dimension and less strongly to neighboring values. Thus, variation along a stimulus dimension, such as wavelength, will, for the most part, be represented by different *sets* of units corresponding to different values on the dimension, rather than the activation level of an individual unit's being the primary indicator of value on the dimension (Thompson, 1965). Figure 1 represents this schematically. Each unit's tuning curve is such that it responds most strongly to a certain value on the dimension, and this response drops off with *distance* from this optimal value. Note that many units will be active when any stimulus on that dimension is present: The coding of position on the dimension is thus in terms of a pattern of activation. Where we are dealing with variations in intensity, we assume that increases in intensity are represented not only by increases in the activity of units already active, but also by the recruitment of additional *neighboring* units. Thus, for both kinds of dimension, the coding of different values on the dimension is achieved partly by differences in which units are activated and partly by differences in their level of activation.

3. Although not random, sampling is selective: Not all the elements of a given stimulus will actually be sampled during presentation, and hence, not all of their corresponding units will be activated on a given trial.

Points 1 and 2 are reasonably straightforward and do not, perhaps, require further explanation at this stage. But Point 3 does. This sampling assumption allows the theory to predict a gradual improvement in performance over successive trials, superimposed on significant variability from one trial to the next. Speed of conditioning to a particular conditioned stimulus (CS) can be directly related both to the absolute number of elements sampled and to the proportion of the total number of possible elements actually sampled on any one trial. Neither of these predictions requires that sampling be purely random, however, and it seems more plausible to suggest that the sampling process should depend on the nature of the stimulus and on any constraints on the organism's ability to inspect or attend to it. For *simple* stimuli, such as tones and colored lights, it is reasonable to suppose that there will be relatively little variability in the sampling from one instant to the next and that a high proportion of elements will be sampled. For more *complex* stimuli, such as a visual shape or pattern, a photograph of a scene, or the experimental context or operant chamber, whose defining characteristics are multiple and distributed over space, the proportion of elements sampled might be expected to be lower and sampling variability higher, owing to the organism's inability to apprehend simultaneously all the features of the stimulus. For example, the perceived odor and texture of an operant chamber will vary from one part of the chamber to another; although the overall geometric shape may be apprehended at once, the finer visual details of one area may require closer inspection of that area to the exclusion of other parts of the chamber. We acknowledge that our model is not completely specified

here and that simulation will, on occasion, require some relatively arbitrary assumptions.

There is, however, another source of variability in sampling that will apply as much to a simple stimulus, such as a tone, as to a more complex stimulus, such as a photograph of a visual scene displayed on a TV screen in a pigeon's operant chamber. In both cases, the organism will sample extraneous elements, arising from other aspects of the environment or from the organism's own behavior or momentary internal state. Since these may fluctuate from trial to trial, they will contribute noise. Since some of these extraneous elements, present and sampled on a given trial, will tend to became associated with the target stimulus (tone or visual scene), their corresponding units will also tend to be activated, even if not sampled, on subsequent presentations of the target.

Fortunately, error-correcting rules (see below) act as signal-averaging algorithms in such circumstances (McClelland & Rumelhart, 1985) and will gradually extinguish these unwanted associations, leading to a less variable representation of the stimulus on each trial. And associative learning will also act to reduce the variability in the representation of a highly complex stimulus, such as a visual scene, whose elements are not all sampled together on a single trial. Over a series of presentations of the picture, the units representing its elements will be activated together more often than the units activated by noise. The units representing the picture will thus become associated with one another more strongly than with any extraneous units, and sampling of any subset of the picture elements will also activate the remainder. In this way, the formation of associations between the elements of a complex stimulus will, in time, reduce the variability in the representation of that stimulus.

This discussion has introduced the idea that associations will be established between any simultaneously activated units. As was noted above, we depart from the restrictive assumption of early versions of stimulus-sampling theory that the only associations formed are between stimuli and responses. Our more liberal assumption allows associations between elements of a target stimulus and a reinforcer or any other stimulus, between stimulus elements and response elements, and between the various elements of a single stimulus.

Association Formation

How are associations formed? The first issue is what associative-learning rule should be adopted—that is, what rule should govern the strengthening and weakening of connections. We follow Rescorla and Wagner (1972) in adopting a rule of the error-correcting class, so that learning is governed by the difference between the input required (in terms of some variable, such as associative strength) and the current input. A *teaching signal* provides external input to a unit (that is to say, input from outside the system, typically from perceptual systems registering a stimulus and specifically not from another unit that may be involved in association formation within the system) and thus specifies the input required to match it. The difference between this external input and the internal input (which is the summed input to the unit from the other units present and available for association) is the error term, Δ, which drives learning. Learning consists of changing associative strength until the internal and external inputs match and the error drops to zero. We note that this class of rule is also favored in modeling human cognitive abilities (e.g., McClelland & Rumelhart, 1985), and it is the delta rule, as specified in that paper, that constitutes the basic learning algorithm in our model. However, we differ in principle (rather than in detail) from Rescorla and Wagner's rule in at least three respects: (1) our treatment of the learning on a given trial, (2) our use of weight decay as a factor controlling learning, and (3) our use of a salience variable to modulate learning to individual units.

Learning on a single trial. All learning in our model is continuous (or a numerical approximation to this). This represents a more realistic approach than the commonly used technique of dividing the learning cycle up into discrete trials. Instead, our simulations can be thought of as having many microtrials within any given conditioning trial. As such, our modeling may be considered to be a closer approximation to real-time simulation. Quite apart from matters of plausibility, this modification has the effect of making it possible to give an account of such phenomena as one-trial overshadowing, which raise problems for Rescorla and Wagner (1972). As we shall argue in a later paper, it also allows us to address some of the phenomena taken by Pearce (1987, 1994) to support configural theories of discrimination learning.

Weight decay. In addition to the basic delta rule, we add the notion of weight decay. Decay represents a mechanism by which transient relationships can be distinguished from stable ones and the latter given preferred status. In our model, each increment to a weight decays exponentially until a fixed proportion of that increment remains, at which point no further decay to that increment ever occurs. As an example, if some learning episode changes a weight between two units by, say, +.32, then if the system were now left entirely to its own devices, that increment would decay in a smooth exponential until it had reached some fixed proportion of its original value, for example, one twentieth. Hence, decay would cease when the increment had been reduced to +.016. At any given time, the total weight of the connection between two units will be composed of many increments, some of which will still be decaying, whereas others will have stabilized and will now be permanent, because they have decayed to asymptote. Thus, any weight can be divided into a part that is decaying and a part that is not, and any weight left long enough would settle to an asymptotic value representing the long-term learning for that connection (see Figure 2).

One effect of this is that learning episodes may now show dissociable short-term and long-term effects. For example, massed learning will typically lead to a short-term advantage (on some appropriate test) over spaced

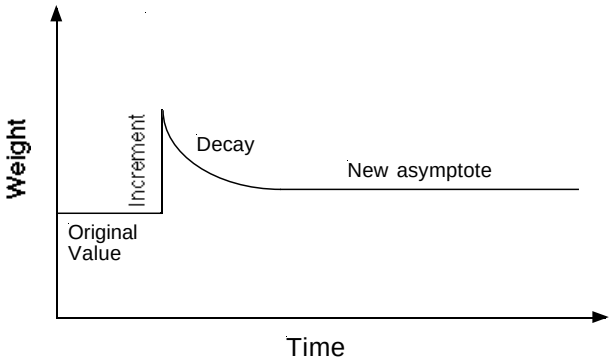

Figure 2. The figure illustrates how the decay function influences learning in the model. If element P is being associated with element Q and the weight depicted is that from P to Q, when P and Q are presented together, the weight is incremented (not instantaneously, but it is assumed that the pairing episode is very brief on the time scale shown) and then decays, so that (in the absence of further presentations) the new final asymptotic value is the old value plus a small proportion of the increment (e.g., one twentieth).

learning, but the opposite will be the case in the long term. As we will note (p. 220), this advantage for spaced practice applies not only to conventional conditioning or associative learning, but also to such paradigms as latent inhibition. Another consequence of the decay mechanism implemented here is that only stable relationships between inputs will ever build up a substantial permanent representation in the system; transient relationships will be quickly learned and then forgotten—an efficient use of the computational resources available.

Salience. The final issue to be tackled here concerns salience modulation. This is accomplished by an additional input to a unit proportional to that unit's error (Δ). This boost is treated by the unit as another component of teaching signal input to be summed with all the other external inputs. Thus, this input will increase both the activation and the error term of this unit. This modulation of external input, however, takes into account how much boost any unit receives and, allowing for this, continues to base its computations on the teaching input (and hence the activation) the unit would have received without the boost. This arrangement has the consequence that learning involving novel unpredicted elements is, other things being equal, faster than that involving familiar (predicted) elements.

This approach is very similar to that taken by Wagner (1978) and has underlying similarities to SOP (Wagner, 1981), although SOP implements the notion of salience modulation as a function of the extent to which a stimulus is predicted in a quite different fashion. One contrast between our model and Wagner's is that we explicitly allow prediction at the elemental level: In other words, the elements of a stimulus can predict one another and thus reduce each other's salience, whereas Wagner tends to consider only associations between the context (or some other stimulus) and a target stimulus as factors influencing the salience of the target. Although both approaches allow for context-specific latent inhibition, we are able to embrace the possibility that a stimulus might display latent inhibition that is not context specific (see p. 224). Salience modulation at an elemental level is also fundamental to our account of perceptual learning.

THE FORMAL MODEL

This concludes our introduction to the model. The following sections give a more formal exposition of the model that allows for quantitative simulation. We take the delta rule (see McClelland & Rumelhart, 1985, for a discussion) as our starting point but later introduce a number of necessary modifications. These modifications are presented step by step, accompanied by a rationale for the alterations made.

The Delta Rule

We first introduce the delta rule and some of the terms and definitions employed in connectionist modeling. Figure 3 shows three elements or nodes interconnected by links or weights. All the nodes are interconnected, but the strength of the link (i.e., the value taken by the weight) between any two nodes may vary. When activation (Ω) tries to pass along a link—that is, when one activated node tries to influence another—the link strength or weight (w) determines how successful this will be. The activation of the emitting node is multiplied by the weight for the link from that node to the recipient; hence, a weight of zero prevents any activation from passing. Note that links are unidirectional but that pairs of nodes are reciprocally interconnected.

In the standard version of the delta rule, the nodes can have an activation of between +1 and −1. This activation is the result of two types of input, external input (e1, e2, e3) and internal input, the latter simply being the input received by a node from other nodes via the links. The summed external input for a given node will be termed

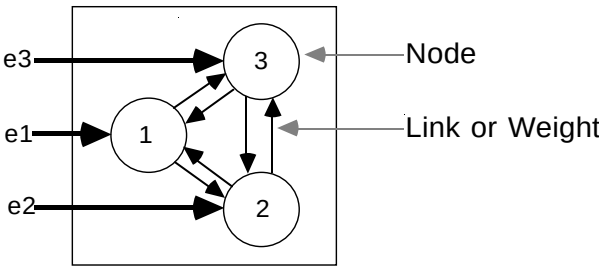

Figure 3. The units and links needed to construct a connectionist system running the delta rule. The system shown is completely connected, and each unit receives external input from outside the system (shown as e1, e2, and e3).

e, and the summed internal input i. The latter is given by Equation 1:

$$i_i = \Sigma\, w_{ij}\, \Omega_j, \tag{1}$$

where i_i is the summed internal input to the ith node, w_{ij} is the weight from node j to node i, and Ω_j is the activation of the jth node.

When input is applied to a node, its activation changes, and when the input ends, the level of activation in the node reverts to its original level. The time course of these changes in activation is described by Equation 2:

$$d\Omega_i/dt = E(e_i + i_i)(1 - \Omega_i) - D\Omega_i, \text{ for } \Omega_i > 0,$$
$$d\Omega_i/dt = E(e_i + i_i)(1 + \Omega_i) - D\Omega_i \quad \text{otherwise}, \tag{2}$$

where E and D are the (positive) constants for excitation and decay, respectively. Broadly speaking, these equations ensure that the level of activation changes until the excitation, $E(e + i)(1 - \Omega)$ for $\Omega > 0$, equals the rate of decay, $D\Omega$, at which point the rate of change of activation is zero and the node is in equilibrium. For example, if e and i are positive and Ω is small, excitation will prevail over decay, and activation will rise, but this decreases the excitation and increases the decay; hence, the rate of increase of activation slows, and activation settles toward some equilibrium level. If the input to the node is now terminated, the decay term smoothly reduces the activation to zero.

However, activation is not the only quantity varying over time; the weights change as well, in a manner that gives the delta rule its name and reveals it to be of the Widrow–Hoff (1960) error-correcting type. Whereas activation is controlled by the *sum* of external and internal inputs to a node, the weights vary in a manner controlled by the *difference* between e and i, the rule being given in Equation 3:

$$dw_{ij}/dt = S(e_i - i_i)\Omega_j, \text{ where } S \text{ is a positive constant.} \tag{3}$$

The term $(e - i)$ is referred to as delta (Δ). The effect of this rule is that, on successive learning trials, the weights into a node are changed until e and i are equal—that is, until the external input is matched by the internal input. Because of the activation term in the rule, it is the weights from the more active nodes that are changed the most.

The account of the delta rule given up to this point has been the standard one (see McClelland & Rumelhart, 1985). Typically, networks employing the rule have been simulated by treating the differential equations given as difference equations, with features such as weight decay and noise treated in an ad hoc manner. The network is allowed to settle to an equilibrium state before any of the weights are changed, thereby dividing processing on a "trial" into discrete stages dealing with activation and, then, learning. In the following sections, we introduce three modifications of the basic rule.

Real-Time Simulation

The first modification is that, in our model, all the processes act on line and continuously; thus, there is no waiting for the system to settle before changing the weights, as is commonly done in other simulations employing the delta rule. We believe that this is a rather more realistic simulation of the learning process than is the standard technique of dividing the learning cycle up into discrete phases. It also, as we shall show, allows us to explain some of the temporal factors governing learning. The way we implement this is to stipulate that quantities such as the external input to a node and the delta for a node must change smoothly rather than abruptly. This involves the system's representing them as activation values themselves; that is, the value of e, the external input to a node, is given by the activation of some hypothetical node or element that responds to the presence or absence of a particular feature of a stimulus. Similarly, Δ or $(e - i)$ is represented by another hypothetical node, whose response to changes in input (representing changes in Δ) is smooth and gradual. The nodes referred to here as hypothetical are only so in that they are not the nodes representing stimuli in the system under consideration but, rather, other bits of computational machinery. Thus no *sharp* discontinuities are introduced into the system, and real-time simulations become possible. Other models (e.g., McClelland & Rumelhart, 1985) have preferred to ignore these complications, at the expense of being unable to track the time course of stimulus processing in the system. Note that the equations given later apply to the values that e and Δ take instantaneously and that, in all the simulations of the model discussed here, the simultaneous differential equations are solved using relatively sophisticated techniques of numerical integration (fourth-order Runge–Kutta).

Decay

Another addition to the basic rule is to add the notion of weight decay. As was mentioned earlier, decay represents a mechanism by which transient relationships can be distinguished from stable ones and the latter given preferred status. At the same time, it is desirable to be able to represent a transient but current contingency between stimuli as well. In order to achieve these somewhat incompatible goals, the idea implemented is that each increment to a weight decays exponentially until a fixed proportion of that increment remains, at which point no further decay to that increment occurs. One consequence of this assumption is that learning episodes may now show dissociable short- and long-term effects. This is because the weight decay process may in itself play a role in controlling learning if the rate of decay is such as to be comparable with the rate at which the delta rule would change the weights, other things being equal. That is to say, the decay mechanism can directly limit learning if the rate of decay comes to nearly balance the rate of increase

of link strength; this will typically occur with massed presentations of input patterns. As a result, the long-term increments that accrue when learning is limited in this fashion are small, but the short-term increment to the weights has to be large for this to be the case.

The discussion of decay given above is, of course, a high-level characterization, rather than a definition. For completeness, the equations governing weight change are given in Equation 4:

$$dw_{ij}/dt = S\Delta_i\Omega_j - Km_{ij},$$

where K is a constant and m_{ij} is a dummy variable.

$$dm_{ij}/dt = S\Delta_i\Omega_j - Lm_{ij},$$

where L is a constant such that $L > K$ ($L, K > 0$). (4)

The relative values of K and L and S determine the balance between short- and long-term increments. The absolute values of K and L will govern how rapidly increments decay to their asymptotic value and, hence, will play a part in determining the relative weight given to short-term learning, as opposed to long-term experience. The use of the dummy variable, m, renders the formulae in a simpler form, avoiding the need for a second-order differential equation.

The Modulation of Salience

Computational principles and algorithmic instantiation. It has been recognized for some time that a problem with many models of associative learning, especially those powerful enough to be interesting, is that they learn very slowly (Rumelhart, Hinton, & Williams, 1986). How can this failing be avoided? There are limits to the performance that can be gained by simply increasing the learning rate parameters (e.g., S) in the equations. In simulation, parameters that are too large can lead to overshoot in learning and oscillation—that is, the changes in the weights can be so large as to exceed those that are required and thus introduce error of the opposite sign into the learning cycle. Even implementation as a genuinely parallel system will not necessarily solve this problem, since the response times of the components may be slow enough to allow overshoot if learning is too rapid. What is needed is a method of enhancing learning that does not suffer from this problem.

A closer analysis reveals one particular source of difficulty. Learning is usually simulated as taking place in a series of steps, each step reducing Δ by some fraction of its current value, with the fraction given by S multiplied by the activation of the inputting element, Ω. Now, Ω normally increases as learning progresses, so that the fraction of Δ taken on each step increases. If the fraction is close to 1 initially, so that early learning is rapid, it will eventually exceed 1 and produce overshoot and oscillation. On the other hand, limiting S to avoid this problem makes early learning unnecessarily slow. The remedy offered here is to modulate the salience of the representational nodes, which in our model will be taken to mean modulating their activity. An alternative would be to let the proposed modulation affect learning only—for example, by having the learning rate parameter S in the delta rule modified on line according to the salience computations. Computationally, the solution via modulation of node activation is the more straightforward, however, and avoids a proliferation of parameters affecting learning and performance.

The idea behind this proposal is that early in learning, when nodes will tend to have large delta values, modulation will be such that the salience of each node will be high and learning will be rapid. As learning progresses, however, modulation will ensure that the activation of the nodes involved falls rather than rises, so that overshoot in the learning never occurs. Another consequence of this approach is that unpredicted elements, represented by nodes possessing a large Δ, will be at an advantage in forming associations, as compared with those representing predicted elements. It might be objected that the rapid learning when elements are unpredicted will result in the system's being at the mercy of coincidental coactivation of nodes. This is true, but the decay mechanism outlined earlier will ensure that no learning episode of this kind will have a permanent effect on the performance of the system.

The particular form of modulation proposed is that the difference between the external and the internal inputs to a node, Δ, should be used to play a part in determining the total external input received by that node: specifically, we assume that the external input will have $r.\Delta$ added to it, where r is a positive constant. This will have two effects. The *effective* error term for a node, Δ', will now become $(r + 1)\Delta$, and the activity of the node will be increased because of the increase in external input. We assume that the system discriminates between Δ and Δ', so that the process of positive feedback does not continue beyond this to generate a Δ'' and so on. A novel combination of elements will, of course, have relatively large error terms, and hence the modulation of external input will greatly increase the error terms and unit activations, promoting association to these nodes (because of the large Δ values) and of these nodes to others (because of their high activation values). It is the latter effect that is properly called salience modulation, whereas the former could be termed modulation of the associativity of the elements and is equivalent to multiplying S by $r + 1$. This salience modulation will have the desired effect on learning, provided that it is sufficient to ensure that, as Δ decreases, the activation of a node decreases as well. In effect, we require that the activation is dominated by the modulation driven by Δ. This means that the fraction by which Δ is reduced on each learning step can now be near 1 initially without causing overshoot, because activation, Ω, will decrease as learning progresses, thereby decreasing the fractional change in Δ. Hence, the network will learn as rapidly as possible early on, when large increments to the associations are possible, but less rapidly later, when Δ is small (see Figure 4).

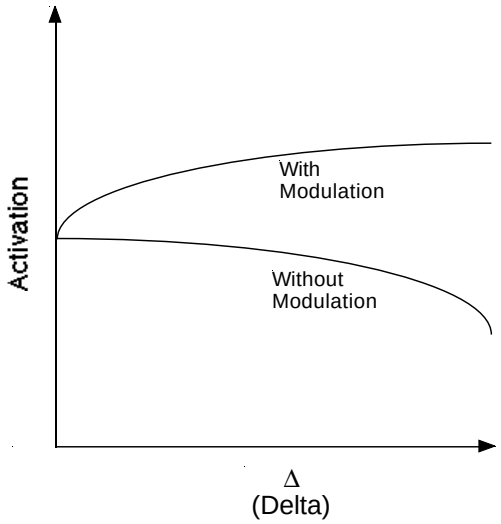

Figure 4. How element activation will vary with Δ, as compared with the unmodulated delta rule. Note how the two curves meet when delta is zero, and how the unmodulated curve starts above zero (when Δ is large) and increases its activation as *i* comes to equal *e*. These are analytic results that hold as long as the parameter *r* exceeds 1 in the modulated case.

Because of the way it enhances learning speed, this modulation implements the heuristic that relatively unpredicted stimuli, represented by nodes with large deltas, will have an advantage in forming associations to other stimulus representations over familiar, predicted stimuli, whose salience will be low. This is reminiscent of the learning rule proposed by Wagner (1978) as a modification of the Rescorla–Wagner (1972) rule. But the application here is somewhat different, and the mechanisms employed to do it are novel.

Implementation. Salience modulation depends on the error Δ—that is, $e - i$—with the proviso that if Δ is negative, no output occurs. As a consequence, (1) nodes activated only internally will never receive any boost, and their activations will be determined entirely by their internal input; (2) a corollary of this is that inhibitory associations (owing to negative correlations between elements) will be formed more slowly than excitatory associations (owing to positive correlations), since the formation of an inhibitory association depends on the absence of a predicted consequence—that is, the node to which inhibitory associations are to be formed being internally rather than externally activated; and (3) excitatory associations will extinguish relatively slowly, since only internally activated nodes are involved.

Summary

The equations governing this system are summarized (with all constants positive) at the bottom of this page. It will be recalled that the boost is added to the external input, *e*, to a given node but that this modulation is taken into account (allowed for) when that node's error term, Δ, is assessed to determine the boost to that node.

A final issue concerns performance in our model, given that we need some mechanism to take us from activations and associative strengths to action. Our approach here is to use the raw output activations of the relevant units to derive measures that will be correlated with performance, while recognizing that although this has the highly desirable property of being transparent to the reader, the reality will be more complex.[1]

This concludes our delineation of the model. We now turn to an analysis of its applications, starting with simple conditioning.

APPLICATION TO SIMPLE CONDITIONING

McLaren et al. (1989) confined their analysis to two contrasting phenomena, latent inhibition and perceptual learning, both of which we will discuss below. But this general associative model should certainly be applicable to other associative phenomena, including simple Pavlovian conditioning. It should be clear how the model outlined above will account for such basic phenomena of conditioning as acquisition and extinction. We have also noted how particular features of conditioning—for example, some of the effects of trial spacing (see Estes, 1955)—can be predicted by the assumption of weight decay. And the sampling assumptions allow, as they do for stimulus-sampling theory, the prediction of trial-to-

$d\Omega_i/dt = E(e_i + i_i)(1 - \Omega_i) - D\Omega_i$	for $\Omega_i \geq 0$, Ω = unit activation
$d\Omega_i/dt = E(e_i + i_i)(1 + \Omega_i) - D\Omega_i$	otherwise
$dw_{ij}/dt = S\Delta_i\Omega_j - Km_{ij}$	where *S* and *K* are constants, m_{ij} is a dummy variable, *w* = weight or connection strength, and Δ is the error term
$dm_{ij}/dt = S\Delta_i\Omega_j - Lm_{ij}$	where *L* is a constant such that $L > K$
$i_i = \Sigma w_{ij}\,\Omega_j$	*i* = internal input to a unit
$\Delta_i = e_i - i_i$	*e* = external input to a unit
modulation = $r\Delta_i$	$\Delta_i > 0$, *r* constant, boost to *i*th node added to *e*
modulation = 0	otherwise (5)

trial variation in performance and a rather natural way of predicting the shape of most acquisition functions.

Excitatory and Inhibitory Conditioning

Excitatory and inhibitory conditioning are dealt with in a reciprocal fashion similar to that of Rescorla and Wagner (1972; Wagner & Rescorla, 1972). Whereas excitatory conditioning is the result of the formation of associative links with positive strengths or weights, inhibitory conditioning is the converse and is instantiated as links with negative associative strength. Thus, input from a link with negative weight will, in the absence of other inputs, produce negative activation of the recipient node. The conditions that result in excitatory or inhibitory conditioning are that (other things being equal) in the former case, there should be a positive correlation between the CS and the unconditioned stimulus (US), and, in the latter, a negative correlation. This simple prescription is somewhat distorted, however, by the fact that the model predicts faster rates of excitatory than of inhibitory conditioning, so that, in practice, a zero correlation between a CS and a US will normally result in modest levels of excitatory conditioning.

McLaren et al. (1989) explained how certain restrictions placed on that earlier version of the model—in particular, the restriction that activations be constrained to be positive or zero—meant that the presentation of a conditioned inhibitor on its own would not result in any extinction of its inhibitory properties, because the positive error term so generated would not be detectable in the network. As has long been clear, the finding that conditioned inhibitors do not extinguish under these circumstances poses serious problems for the original Rescorla–Wagner model (Baker, 1974; Zimmer-Hart & Rescorla, 1974). But there are now reasons to believe that this restriction will create other problems for our model. In particular, the results of experiments by Espinet, Iraola, Bennett, and Mackintosh (1995) and Bennett, Scahill, Griffiths, and Mackintosh (1999), discussed later in more detail (p. 235), suggest that units representing a CS may take on negative activation. Our present solution to this problem is to allow representational units to have negative activations, but not to allow negative internal input to be detectable by the mechanism responsible for computing error. This would ensure that inhibitors would not extinguish but still allow us to explain the results of Espinet et al. (1995).

Selective Associations

The incorporation of error correction, in the form of a modified delta rule, predicts the phenomena of overshadowing, blocking, and contingency effects, now taken as the touchstone of any satisfactory theory of associative learning. The explanation is the same as that provided by Rescorla and Wagner (1972): Weight changes between CS and US units (changes in the associative strength of a CS) are proportional to an error term that is a function of the discrepancy between external and internal inputs to US units ($\lambda - \Sigma V$, in Rescorla and Wagner's terminology). The most straightforward application is to blocking, where conditioning to CS_2, reinforced in compound with CS_1, is reduced by prior conditioning to CS_1: The prior conditioning to CS_1 reduces the value of the error term on conditioning trials to CS_2. Overshadowing, where conditioning to CS_2 is attenuated if it is conditioned in compound with a more salient CS_1, is assimilated to blocking—as it is by Rescorla and Wagner: The more rapid conditioning to the more salient CS_1 rapidly reduces the error term for conditioning to CS_2. Rescorla and Wagner, it should be noted, have a certain problem in predicting overshadowing on the first trial of conditioning (J. H. James & Wagner, 1980; Mackintosh & Reese, 1979). Given no prior conditioning to either CS_1 or CS_2, on Trial 1 the error term ($\lambda - \Sigma V$) will be unaffected by the presence or absence of CS_1. Our method of approximating to real-time processing, by effectively dividing up individual trials into a series of microtrials, allows competition for associative strength to develop during the course of the first conditioning trial and thus avoids this problem.

Generalization and Discrimination

After conditioning to a particular CS—say, a 1-kHz tone signaling the delivery of food—animals will respond not only to the original CS, but also to tones of other frequencies or of the same frequency but different amplitude. We follow stimulus-sampling theory in assuming that such generalization occurs because similar stimuli share elements in common. Two tones, differing in frequency, can be conceptualized as overlapping sets of elements, AX and BX, where A and B are the elements unique to each and X are the elements common to both. Following conditioning to AX, responding will generalize to BX because conditioning involved changing the associations of the X nodes as well as those of the A nodes.

Discriminative training (e.g., reinforcement of responding to AX and nonreinforcement of responding to BX) reduces generalization of responding from one stimulus to the other. Our analysis of this follows that of Rescorla and Wagner (1972). If AX is reinforced and BX is not, error correction will ensure that excitatory conditioning will accrue to A elements and inhibitory conditioning to B elements, whereas the common X elements, being relatively poor predictors of the outcome of each trial, are overshadowed (see also the analysis of Wagner, Logan, Haberlandt, & Price, 1968, given by Rescorla & Wagner, 1972). Since generalization from AX to BX depends on the conditioning of these X elements, it will be reduced. Pearce (1987, 1994) has pointed to a number of problems for such an elemental analysis of discrimination learning. We discuss them in detail in the forthcoming companion paper to this one.

Temporal Characteristics of Conditioning

The approximation to a real-time system allows the model to account for some of the temporal characteris-

tics of conditioning, because the pattern of CS activation associated with a US is that generated by the CS at the moment of US presentation. Given the latency for CS activation to rise to a maximum and given that this latency does not apply to the error term governing learning, we can account for the finding that, for maximal conditioned responding, the CS should precede the US and can also predict that the conditioned response (CR) should anticipate the US. On this analysis, many of the phenomena explained by Miller in terms of a *temporal coding hypothesis* (Blaisdell, Denniston, & Miller, 1998; Cole, Barnet, & Miller, 1995) are explained on the assumption that the pattern of activity generated by the CS at the moment of US onset differs from that generated by the CS at other times (e.g., by its onset). This is, of course, the explanation offered by Hull (1943).

Consider, for example, the results of a sensory preconditioning experiment reported by Cole et al. (1995). In the first stage of the experiment, rats received paired presentations of two 5-sec CSs, CS_1 and CS_2, with CS_2 immediately following CS_1. In the second phase, CS_1 signaled the delivery of a 0.5-sec shock, with the shock being delivered either immediately (Group 0) or 5 sec (Group 5) after the offset of the CS. In the final phase, both groups were tested for suppression to CS_2. As one would expect, there was more suppression to CS_1 when it was immediately paired with the delivery of shock in Group 0 than when there was a 5-sec trace interval between CS_1 and the US. But this difference was reversed in the case of CS_2: There was more suppression to CS_2 in Group 5 than in Group 0. Cole et al. argued that this finding is paradoxical, since, other things being equal, one would expect the CR to a sensory preconditioned CS_2 to depend on the strength of conditioning to the CS_1 with which it had been paired. But other things are not equal. Let us assume that sensory conditioning depends, at least in part, on the establishment of an association between the representation of CS_2 retrieved by CS_1 and the US. Since CS_2 followed CS_1, that representation will be maximally active after the termination of CS_1. It follows that for Group 0, this activation will peak after US presentation, whereas for Group 5, it will precede US presentation. Since forward pairings are more effective than backward pairings, one would expect to see more suppression to CS_2 in Group 5 than in Group 0.

Much the same analysis applies to a second-order conditioning experiment, of very similar design, also reported by Cole et al. (1995). Here, the only problem for our model, as it is for Wagner's SOP, is to explain how an excitatory, rather than an inhibitory, association is formed between the representation of CS_2 and the retrieved representation of the US. But as Sutton and Barto (1981) make clear, our type of approach may not be sufficient to explain all the temporal characteristics of conditioning and, in particular, why there should be an optimal interval between CS and US for successful conditioning that differs for different conditioning preparations (see Mackintosh, 1983).

The idea of weight decay provides a second way in which our model addresses certain temporal factors in conditioning—in particular, the phenomenon of spontaneous recovery. Following a series of nonreinforced trials, sufficient to extinguish responding to a previously reinforced CS, the mere passage of time is sufficient to restore a significant level of responding to that CS (Mackintosh, 1974). As Wagner and Rescorla (1972) noted, the occurrence of spontaneous recovery presents something of a problem for a theory that sees excitatory and inhibitory conditioning simply as changes in a single underlying variable of associative strength. Our explanation makes use of the fact that inhibitory increments to an associative link are just as susceptible to our postulated decay mechanism as are excitatory increments. In this case, the association may well have been reduced to near zero by the extinction phase, but with time the inhibitory increments decay, and that allows the association to increase in strength to a level where it can once again support conditioned responding. Thus spontaneous recovery can be understood in terms of weight decay in a real-time model.

Retrospective Revaluation and Mediated Conditioning

There is nothing in our model that requires external input to CS units in order for that CS to be associated with some further consequence. Indeed, we have already suggested that one of the effects of the formation of associations between the various elements of a CS is that those elements actually sampled on a given trial will activate units corresponding to unsampled elements, thus allowing changes in the weights of their connections to US units. In effect, this is quite inconsistent with the original version of SOP (Wagner, 1981), where the associative strength of a CS is assumed to change only on trials when the CS is present. According to SOP, the representation of a stimulus must be in one of three states: primary activity, labeled A_1; secondary activity, labeled A_2; and inactivity (I). The presentation of a CS drives its representation into A_1, from which it decays into A_2 before finally decaying into I. The presentation of an associate of that CS—for example, the context in which it has previously occurred—retrieves the representation of the CS directly into A_2. The associative rules of SOP are that an excitatory association is established between a CS and US only when they are both in A_1. If the CS is in A_1 and the US is retrieved into A_2, an inhibitory association is established between the CS and the US. A CS in A_2 enters into no new associations, excitatory or inhibitory, with a US.

Unlike SOP in its original form, if internal activation of a node is sufficient to allow the formation of associations between that node and others, our model is able to account for *mediated conditioning* (Holland, 1981). In the first phase of one of Holland's experiments, a tone signaled the availability of sucrose pellets; in the second phase, the tone signaled an injection of lithium chloride.

The combination of these two treatments was sufficient to establish an aversion to the sucrose pellets, presumably because, although their actual consumption was never paired with lithium, the prior association between tone and sucrose pellets ensured that an associatively retrieved representation of them was. In the terminology of SOP, an excitatory association was established between a CS in A_2 and a US in A_1.

Although such instances of mediated conditioning (see also Ward-Robinson & Hall, 1996) pose no problem for our model, evidence of retrospective revaluation does. In the first phase of such an experiment, a CS_1 + CS_2 compound signals the delivery of a US. In the second phase, CS_1 is presented alone for a series of treatment trials, and the effect of this treatment on responding to CS_2 is evaluated in a final test phase. Evidence of *backward blocking* is provided if reinforcement of CS_1 in Phase 2 attenuates evidence of conditioning to CS_2 (Miller & Matute, 1996); evidence of *unovershadowing* is provided if extinction of CS_1 increases evidence of conditioning to CS_2 (Matzel, Schactman, & Miller, 1985). One interpretation of this evidence of apparent retrospective revaluation of CS_2 is in terms of Miller's comparator hypothesis, which attributes the effect to a change in *performance* to CS_2, rather than to any change in learning about CS_2 (Miller & Matzel, 1988). But an alternative interpretation (Dickinson & Burke, 1996; Van Hamme & Wasserman, 1994) suggests that changes to the associative strength of the absent CS_2 do occur during the course of Phase 2 because its representation is retrieved into memory (in terms of SOP, into A_2) by the presentation of its associate, CS_1. However, these associative changes must be opposite in sign to those inferred in experiments on mediated conditioning. If pairing CS_1 with the outcome in Phase 2 of a backward blocking experiment reduces the associative strength of CS_2, this implies that the pairing of CS_2, retrieved into A_2, with the US in A_1 results in extinction of CS_2. It must equally be assumed that unovershadowing occurs because presentation of CS_2 alone in the second phase of the experiment retrieves both CS_1 and the US into A_2 and this causes an increase in the strength of the association between the two.

Direct evidence from flavor preference experiments (Dwyer, Mackintosh, & Boakes, 1998) has indeed shown that the associatively activated representation of a flavor can apparently change its associative strength in exactly the way implied by Dickinson and Burke (1996) and Van Hamme and Wasserman (1994). Dwyer et al. found that internal activation of CS units coincident with internal activation of US units can sometimes increase the weight of their connections: In effect, they found evidence of excitatory conditioning to a CS when both the CS and the US were in the A_2 state. We offer no explanation for such observations now; we note only that they imply the operation of associative rules that are diametrically opposed to those required to explain mediated conditioning (but see Dwyer, 1999).

APPLICATION TO LATENT INHIBITION

The Analysis of Latent Inhibition as a Reduction in Salience

We assume that the salience or associability of any stimulus is affected by the operation of salience modulation. As was noted earlier, this modulation provides an additional input to the activation of a unit, directly proportional to the discrepancy, Δ, between all other external and internal inputs to that unit. The effect of this is similar to that achieved in SOP (Wagner, l981) by the assumption that a CS, all of whose elements have been retrieved into the A_2 state by the presentation of an associate of that CS, will not enter into any new associations on that trial. Although the actual processes that cause this loss of associability are quite different in SOP from that envisaged here, the consequences are very much the same. All the stimuli start with high associability, but as their occurrence becomes expected, so their associability declines. In both theories, CSs are conceptualized as sets of elements, and it is the associabilities of individual elements that decline. The most important difference is that SOP assumes that the units activated by a CS can be retrieved into the A_2 state only by the presentation of other associates of the CS. Since we assume that the repeated presentation of a CS results in the formation of associations between all its elements, one source of internal input to any unit activated by a CS will be from other units also activated by the CS. Put informally, SOP states that a CS will lose associability to the extent that its occurrence becomes expected (say, because it has been repeatedly presented in this context); although accepting that this is one reason why a CS loses associability, we are also saying that another reason can be simply that the stimulus becomes familiar.

It is worth noting that a further assumption that we share with SOP is that latent inhibition, defined as retarded acquisition of a CR owing to CS preexposure, reflects a loss of associability. Preexposure to a CS reduces the ability of that CS to enter into new associations. We do not view latent inhibition as a performance effect or as a failure to retrieve a new CS–US association as a result of competition from a previous CS–nothing association (see, e.g., Bouton, 1993: Hall, 1991; Miller, Kasprow, & Schactman, 1986). We do, however, acknowledge that there is evidence consistent with such alternative accounts.

An Illustration: Massed Versus Spaced Stimulus Preexposure

Here, we illustrate the application of the model to stimulus preexposure with an example that incorporates all three of the novel features that distinguish it from the basic delta rule: decay, modulation, and real-time pro-

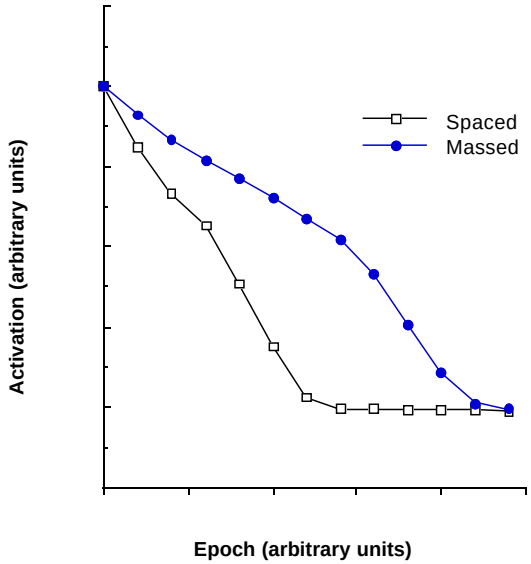

Figure 5. Simulation of massed versus spaced preexposure to stimulus elements and its effect on element salience. The activations shown are after weight decay has been allowed to take place, simulating the effect of preexposure on one day and then testing on the next. Note how for intermediate levels of preexposure, the long-term latent inhibition accruing in the spaced schedule is substantially greater than that in the massed case.

cessing. The example we shall consider is that of massed versus spaced preexposure to a stimulus and its consequences for the loss of salience, contingent on preexposure, predicted by the model. We note that there is evidence that latent inhibition will be stronger, other things being equal, after spaced preexposure than after massed preexposure to a stimulus (Lantz, 1973; Schnur & Lubow, 1976), and this is what the model predicts by virtue of its decay function in conjunction with modulation. Figure 5 confirms this by plotting element activation as a function of amount of either massed or spaced exposure to a stimulus. The spaced schedule is more effective in promoting durable associations among elements, which in turn leads to strong modulation of element activity so that the salience of any stimulus containing these elements would be reduced. This would result in strong latent inhibition to the preexposed stimulus containing these elements.

Elemental Versus Configural Theories of Latent Inhibition

An elemental theory states that preexposure to any complex stimulus causes latent inhibition of the elements of which that stimulus is composed, rather than of the stimulus as a configural whole. By the same token, preexposure to one component of a compound stimulus causes latent inhibition of that component alone, which may well lead to superior conditioning to the other component of the compound when the compound CS is paired with a reinforcer. One line of evidence consistent with this analysis comes from a number of our studies of perceptual learning in flavor aversion conditioning. For example, an injection of lithium following rats' consumption of a novel saline–lemon solution conditions an aversion to saline–lemon that generalizes strongly to a sucrose–lemon solution (Mackintosh, Kaye, & Bennett, 1991). Unreinforced preexposure to lemon alone has relatively little effect on the conditioning of the aversion to saline–lemon but sharply reduces its generalization to sucrose–lemon. We attribute the normal generalization of the aversion from saline–lemon to sucrose–lemon to the conditioning of that aversion to the common lemon flavor. The reduction in generalization is thus most plausibly attributed to the effect of preexposure on the associability of lemon. A variety of other experiments have confirmed that latent inhibition of the element or elements common to two or more stimuli will significantly reduce generalization between them (see p. 226).

More direct tests of elemental theory's expectations here are provided by experiments by Carr (1974) and Navarro, Hallam, Matzel, and Miller (1989). In both studies, rats received conditioned suppression training, in which shock was predicted by a compound CS (AB) or by B alone. The relative intensities of A and B were such that A overshadowed B: Animals conditioned to the AB compound showed significantly less suppression to B than those conditioned to B alone. But when animals were preexposed to A, this overshadowing effect was abolished. Latent inhibition of one component, A, of the compound CS actually enhanced the level of conditioning to the other component, B. These results have been replicated by Darby and Pearce (1997) in studies of appetitive conditioning: Rats received concurrent conditioning trials to two compound CSs—AB and AC—and were tested for their level of responding to B and C alone. Latent inhibition of A significantly increased animals' level of responding to B and C.

On one reading at least (as was acknowledged in Darby & Pearce, 1997), configural theory predicts a quite different result. If latent inhibition is treated like conditioning—as is explicitly envisaged by, say, Bouton (1993) and Hall (1991)—then, according to a configural analysis, latent inhibition of A should generalize to the AB compound, retarding conditioning to the compound when it is paired with a US. Moreover, since responding to B on test is dependent on generalization from AB, any reduction in the level of conditioning to AB should also lead to a reduction in responding to B. Thus latent inhibition of A, so far from increasing responding to B, should, if anything, decrease it.

Darby and Pearce (1997) have shown that it is possible for configural theory to escape this dilemma by assuming that latent inhibition is simply a matter of a reduction in the salience of a preexposed stimulus. In Pearce's (1987, 1994) theory, generalization from a compound CS (AB) to its components (A or B) is a function of the proportion of elements A and B share with the compound, and that will depend on the relative salience of A and B. Reason-

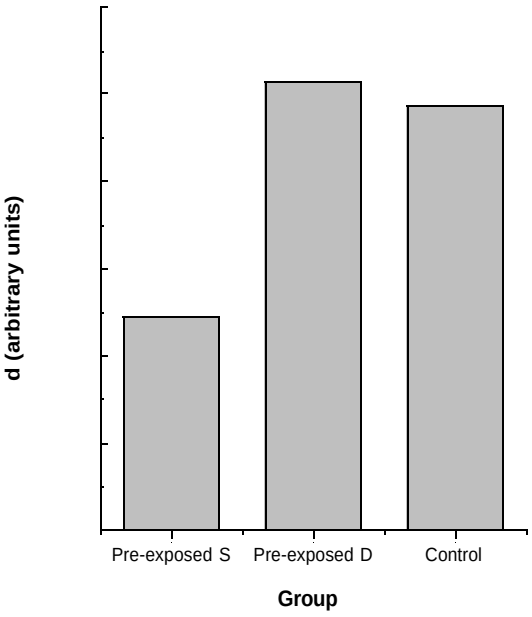

Figure 6. Simulation of the context specificity of latent inhibition. In this simulation, there were three groups. The Preexposed S group was preexposed to the conditioned stimulus (CS) and conditioned in the same context, the Preexposed D group was conditioned in a different (though equally familiar) context to that of stimulus preexposure, and the controls were simply preexposed to the context alone and then conditioned to the CS in that context.

ably enough, Pearce states that if A is of low salience and B is of high salience, there will be substantially more generalization from AB to B than from AB to A. It follows that if A and B are, intrinsically, of equal salience but preexposure to A reduces A's salience, such preexposure will *increase* generalization from AB to B. Hence, preexposure to one element of a compound CS may increase the evidence of conditioning to the other component. It remains to be seen whether this analysis provides a more satisfactory account than the simple, elemental analysis. What seems certain is that it will have some difficulty in handling the various effects of context on latent inhibition, which we turn to next.

Context Sensitivity of Latent Inhibition

One of the best established facts about latent inhibition is its context sensitivity. If preexposure and conditioning to a CS take place in the same context, preexposure will retard conditioning, but this latent inhibition effect is attenuated, and sometimes even abolished, if preexposure and conditioning take place in different contexts (see Hall, 1991, for a review). Although early demonstrations of this effect confounded a change of context with the absolute novelty of the conditioning context, later experiments have made it clear that the critical factor is the change of context between preexposure and conditioning. Lovibond, Preston, and Mackintosh (1984), for example, exposed rats to two different CSs in two different contexts, A in context C_1 and B in context C_2. Group Same then received conditioning trials to A in C_1 and to B in C_2, whereas Group Different was conditioned to A in C_2 and to B in C_1. Group Same showed a significantly greater latent inhibition effect than Group Different.

This basic finding is predicted by SOP and has long been thought to provide good evidence in favor of Wagner's analysis of latent inhibition (although it is also predicted by Miller & Matzel's, 1988, comparator hypothesis). According to SOP, the effectiveness of any conditioning episode is a function of the proportion of CS and US elements that are simultaneously in the A_1 state. As a result of preexposure to a CS in a given context, associations will be established between that context and that CS, with the consequence that on subsequent conditioning trials in that context, contextual cues will retrieve a representation of the CS into A_2 and conditioning will be impaired. If conditioning occurs in a discriminably different, even if equally familiar, context, this different context will not retrieve the CS into A_2, and conditioning will proceed rapidly. Our own assumption that the salience of a CS is a function of the discrepancy between external and internal inputs to the units activated by the CS yields exactly the same prediction. Sufficient exposure to a CS in a given context will result in an increase in the weights of the connections between units activated by the context and those activated by the CS, and on subsequent presentations of that CS in that context, the discrepancy between external and internal inputs to the CS units will be reduced (see Figure 6 for a simulation of the context specificity of latent inhibition).

Our account departs from anything that Wagner has explicitly stated in allowing that there are other sources of latent inhibition. Since we conceptualize all stimuli as sets of elements and allow the formation of associations between the elements of a single stimulus, such intra-CS associations will act to reduce the discrepancy between external and internal inputs to individual units equally and, thus, reduce the salience of the elements comprising the CS. According to this analysis, therefore, the extent to which latent inhibition is disrupted by a change of context will depend not only on such obvious factors as the discriminability of the two contexts, but also on the relative strengths of the context–CS associations and of the intra-CS associations produced by prior exposure to the CS in a given context. Since the formation of these associations will follow our standard error-correcting rule, the two sets of associations will be in competition with one another. This allows us to predict that, in certain circumstances, latent inhibition will show little or no disruption with a change of context between preexposure and conditioning (e.g., after context preexposure, see p. 224).

On this analysis, it should also be possible to obtain context-independent latent inhibition by the use of a CS that permits multiple within-CS associations, which would compete more effectively with the formation of context–CS associations. In an unpublished set of experiments, Plaisted and Mackintosh preexposed and then

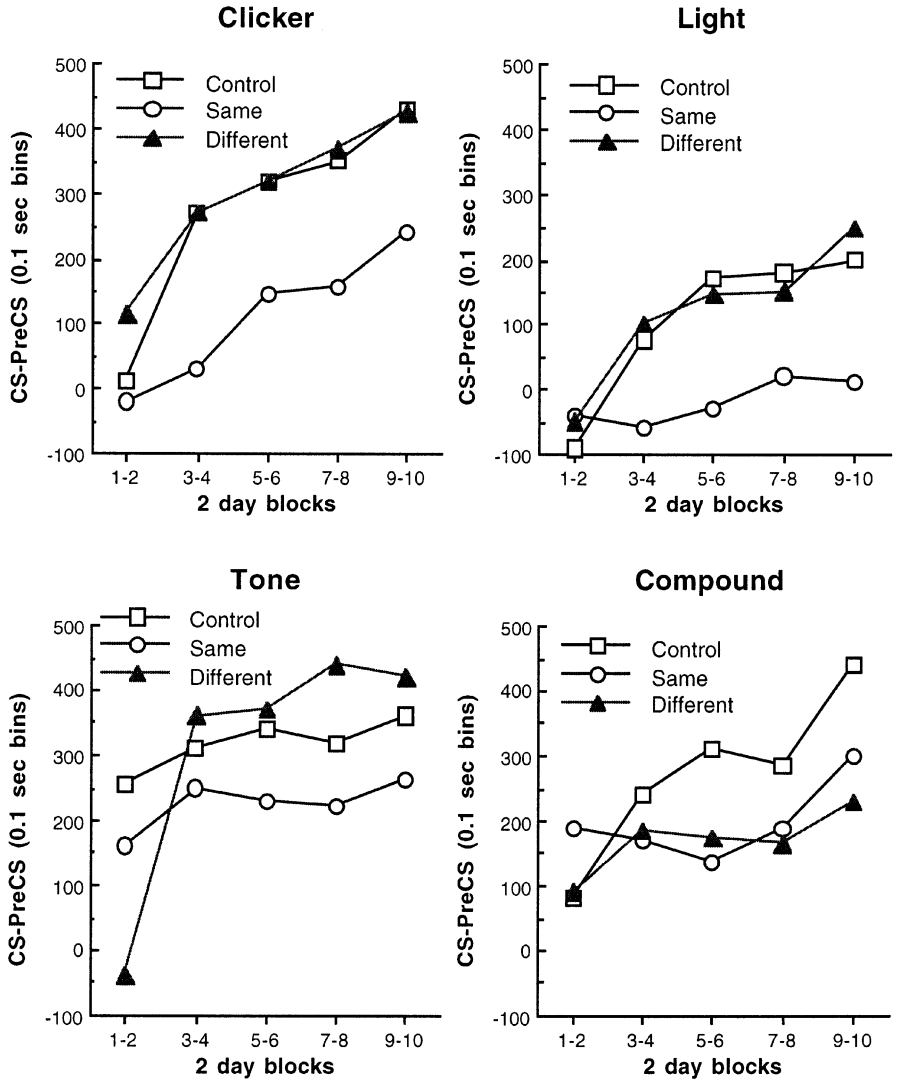

Figure 7. Context sensitivity of latent inhibition when the conditioned stimulus (CS) is either an elemental stimulus or a compound (Plaisted & Mackintosh, unpublished). Each panel shows the acquisition of conditioned responding (magazine approach) by three groups of thirsty rats to a 10-sec CS signaling the delivery of water. The CS was either a 2-kHz tone (bottom left panel), a 1.5-Hz clicker, (top left), a pair of flashing (0.5-Hz) keylights (top right), or a compound of all three (lower right). Within each panel, Group Same had received 40 trials of preexposure to their CS in the context in which conditioning later took place, Group Different received 40 trials of preexposure to the CS in a different context (but received equal exposure to the conditioning context), and the control group received exposure to both contexts with no CSs presented. The contexts were identical operant chambers, differing in odor and the time of day at which sessions were run (morning and afternoon). For all four CSs, there was a significant effect of group on conditioning; in the case of each of the elemental CSs, this difference was due to slower conditioning in Group Same than in the other two groups, which did not differ significantly from one another. In the case of the compound CS, Groups Same and Different did not differ significantly from one another, but both were significantly slower in acquisition than the control group.

conditioned thirsty rats to a compound CS consisting of the simultaneous presentation of a tone, a clicker, and a flashing light or to one of these elements alone. Preexposure and conditioning took place either in the same context or in different contexts—although in the latter case, the conditioning context was equally familiar to the rats. Control groups received no preexposure to the CS before conditioning, although for these animals also, the conditioning context was as familiar as it was for preexposed animals. The results of the conditioning phase are shown in the four panels of Figure 7. The first three panels show the results for the three sets of groups, each conditioned

to one of the three elements. The results show the expected latent inhibition effect in animals preexposed and conditioned in the same context (Same) and its expected disruption when animals were preexposed in one context and conditioned in another (Diff). The final panel shows the results for the group conditioned to the tone–clicker–light compound. It is evident that both Groups Same and Diff show an equally strong latent inhibition effect. We interpret this result to mean that the greater the opportunity for the development of within-CS associations produced by the use of a compound CS, the more such associations will prevent the formation of context–CS associations, and the more the latent inhibition effect observed will transfer from one context to another.

Context Preexposure

If the formation of associations between the various elements of a complex stimulus leads to a reduction in the salience of the elements of that stimulus, preexposure to an experimental context should reduce the salience of its elements or features, which will then be slow to enter into new associations. Context preexposure should result in latent inhibition of the context. If a CS is then preexposed in that context, the latent inhibition of the context will retard the formation of context–CS associations. To the extent that the formation of such associations competes with the formation of within-CS associations, we should expect context preexposure to enhance the formation of within-CS associations. Although this may or may not lead to some overall reduction in the magnitude of latent inhibition to the CS, what it will certainly mean is that any such latent inhibition should transfer to another context. McLaren, Bennett, Plaisted, Aitken, and Mackintosh (1994) confirmed this prediction. In one experiment, this abolition of the context specificity of latent inhibition was accompanied by an apparent decrease in the overall magnitude of latent inhibition, but in another it was not.

These results also pose something of a problem for theories that seek to explain latent inhibition as a case of associative interference in which preexposure to a CS establishes a CS–nothing association that then interferes with the formation of a CS–US association during subsequent conditioning trials or with the retrieval of this association on a subsequent test trial (Bouton, 1993). If latent inhibition is the consequence of these sorts of associations, it should be subject to blocking. Prior establishment of a context–nothing association should block the formation of the CS–nothing association when the CS is preexposed in that context. The second of McLaren, Bennett, et al.'s (1994) experiments provided no evidence at all of any such reduction in latent inhibition, and Hall and Channell (1986) actually reported that preexposure to the context increased the magnitude of latent inhibition to a CS exposed in that context.

Context Extinction

We follow SOP in assuming that the context sensitivity of latent inhibition is a consequence of the establishment of context–CS associations. From this it would seem to follow that, where latent inhibition is disrupted by a change of context between preexposure and conditioning, it will be equally disrupted by the interpolation of some sessions of exposure to the context alone between preexposure to the CS in that context and subsequent conditioning to the CS in that context. The interpolated exposure to the context alone should extinguish the context–CS associations that are assumed to be largely responsible for latent inhibition in the first place.

The effect of such context extinction on latent inhibition has been examined in a number of studies. Hall and Minor (1984), for example, reported six experiments on conditioned suppression in rats that uniformly failed to find any reduction in latent inhibition as a result of such treatment. Baker and Mercier (1982) also reported six conditioned suppression experiments—only three of which found evidence of any reduction in latent inhibition. The main difference between those experiments in which context extinction had an effect on latent inhibition and those in which it did not was that, in the latter case, conditioning trials were only partially, not consistently, reinforced. (But this is not a sufficient explanation of Hall and Minor's negative results.) Wagner (1979) has also briefly reported a study of conditioned suppression in which a reliable effect of context extinction was found, and other conditioning preparations have been more consistently successful. Wagner (1979) reported a second study, of rabbit eyelid conditioning, in which context extinction reduced latent inhibition. And in a study of conditioned odor aversion, Westbrook, Bond, and Feyer (1981) found that latent inhibition was abolished by exposure to the context alone between CS preexposure and conditioning (see also Grahame, Barnet, Gunther, & Miller, 1994).

We do not pretend to have a full or adequate explanation for these conflicting results. It is simply not clear why exposure to the context alone after preexposure to a CS in that context should sometimes attenuate or even abolish latent inhibition and, at other times, have essentially no effect. We should stress, however, that our own account of latent inhibition, although certainly leading one to expect some attenuation of latent inhibition, especially in circumstances in which latent inhibition shows context sensitivity, does not predict that latent inhibition will be abolished by such a treatment. That part of the effect that is attributable to within-CS associations should survive the extinction of context–CS associations. The one prediction that does follow from our account is that anything that increases the relative importance of within-CS, as opposed to context–CS, associations should decrease the effectiveness of a context extinction treatment. The discussion in the preceding section suggests some obvious ways of testing this prediction. It is, perhaps, worth noting that the experiments that have failed to detect an effect of context extinction have employed conditioned suppression, which is a procedure that typically involves significant pretraining of a baseline response (i.e., significant prior exposure to the context).

APPLICATION TO PERCEPTUAL LEARNING: THEORY

One of the main applications of our earlier (McLaren et al., 1989) formulation of the model was to the phenomenon of perceptual learning. Ten years on, we still believe that this application has been successful and that we have provided a more clearly articulated, more powerful, and more testable account of perceptual learning than has hitherto been available.

The Phenomena to be Explained

We should, at this point, define our terms carefully. It is, of course, a common observation, and one readily documented in the laboratory, that prolonged practice, especially instructed practice, will enhance people's ability to discriminate stimuli that were initially as indistinguishable to them as they still are to others. William James (1890) provided a number of examples of the remarkable feats of expert tasters and testers of various products. More recent experiments on *hyperacuity* have shown that human observers can, after sufficient experience, make accurate vernier acuity judgments of displacements as small as 5–10 sec of arc—although the diameter of foveal cones is about 30 sec of arc (Poggio, Fahle, & Edelman, 1992).

Any worthwhile theory of learning should have no difficulty in predicting that prolonged instructed practice, or differential reinforcement, will result in successful discrimination between two or more stimuli that the observer initially reacted to in the same way. The theoretical challenge is to explain how mere exposure to two or more stimuli, in the absence of differential reinforcement, should enhance an observer's ability to discriminate between them. This is the phenomenon of perceptual learning we wish to address. We must acknowledge, however, that the experimental challenge, when dealing with human subjects, is to ensure that the exposure is, indeed, "mere"—that is, that there has been no implicit instruction or differential reinforcement. If experimental subjects are shown a series of stimuli, with the instructions being simply to look at them or study them, but without trying to discriminate between them, they will treat such instructions with the contempt that they deserve. How can they study some stimuli without looking for differences between them? And why should they? The chances are that a later stage of the experiment will precisely require them to notice such differences. Although much is made of the absence of explicit feedback in many experiments on perceptual learning in people, as in those on hyperacuity, the fact remains that observers are told that they will be shown a series of stimuli that differ in certain specified ways (e.g., that one line is to the left or right of another) and that their task is to say which. Even if they are never told whether their answers are right or wrong, it is not difficult to see how they might learn that what initially looked like random noise in the stimulus array would in fact provide the clue to the correct answer.

This is in sharp contrast to the situation in a typical animal experiment on perceptual learning. In E. J. Gibson and Walk's (1956) classic demonstration, rats trained on a discrimination between a circle and a triangle learned very much more rapidly if they had lived for the past month in cages with circles and triangles hanging from the walls than if the stimuli were wholly novel. E. J. Gibson and Walk's experiment was followed by a number of other, similar studies documenting that home cage exposure to various visual stimuli would facilitate subsequent discrimination learning (see Hall, 1980, for a review). A quite different procedure has been more commonly used in recent studies. Thirsty rats are given a flavored solution to drink, followed by an injection of lithium chloride. The aversion conditioned to this flavor will then generalize to another, similar flavor. Perceptual learning is demonstrated by the observation that rats given prior exposure to the two flavors, by drinking measured amounts of each over a period of several days, will show significantly less generalization of the aversion from one flavor to the other than rats that have not received such prior exposure (e.g., Honey & Hall, 1989; Mackintosh et al., 1991). In another procedure we have employed, the exposure phase involves placing rats on the arms of a radial maze or on a small platform in the center of a circular pool. Such exposure will facilitate the subsequent learning of a spatial discrimination between two arms of the maze or the rats' ability to find a submerged platform in one quadrant of the pool (see, e.g., Chamizo & Mackintosh, 1989; Prados, Chamizo, & Mackintosh, 1999; Rodrigo, Chamizo, McLaren, & Mackintosh, 1994).

In none of these experiments did the exposure phase involve any explicit differential reinforcement for attending to differences between the various stimuli, let alone for responding differently to them. The circles and triangle hanging on the walls of the rats' cages in E. J. Gibson and Walk's (1956) experiment signaled nothing of consequence, and it seems most likely that the rats would soon have simply ignored them. The dilute solutions given to thirsty rats will have tasted slightly different, but nothing hung on this difference. The important feature of the solutions was presumably that they quenched the rats' thirst. So why should such exposure have any effect on animals' subsequent ability to discriminate between these stimuli?

Three Mechanisms for Perceptual Learning

Here, we illustrate the model by simulating three mechanisms for perceptual learning. Each simulation illustrates the operation of one of the associatively based mechanisms for perceptual learning discussed earlier by McLaren et al. (1989). We start by considering the case in which latent inhibition can result in faster learning.

Latent inhibition of common elements. Take two stimuli moderately similar to one another, which we label AX and BX. In terms of the representations used in this example, the state of affairs will be as shown in Figure 8.

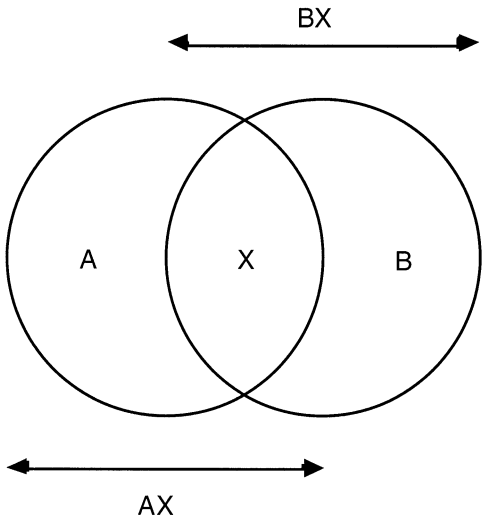

Figure 8. The representation of two similar stimuli, AX and BX, in elemental terms. See the text for a discussion.

Both stimuli have unique elements (A and B elements), but they share the X elements in common: These X elements are the basis of their similarity and would also be the basis for any generalization between them. If BX is preexposed for some time before AX is paired with a US, less conditioning should generalize to BX, as compared with a control group that received no preexposure. This is because the X elements will be latently inhibited (and so have reduced salience) by preexposure; they will therefore be overshadowed by the A elements, which will acquire most of the associative strength to the US, leaving less strength to accrue to the X elements and hence generalize to BX. Honey and Hall (1989) have performed this experiment with flavors (see also Bennett, Wills, Wells, & Mackintosh, 1994). This type of study qualifies as a demonstration of perceptual learning because simple stimulus exposure leads to a greater refinement of stimulus representation, as measured by the generalization test.

Even when animals are preexposed to both AX and BX, latent inhibition may facilitate their subsequent discrimination. Ten trials of preexposure to AX and 10 to BX implies 10 trials of preexposure to both A and B elements but 20 trials of preexposure to X. If latent inhibition is some increasing function of the amount of preexposure, the salience of the X elements will have been reduced more than that of the A and B elements, and this will, other things being equal, result in more rapid discrimination between the two stimuli.

A simulation of this mechanism is illustrated in Figure 9. The simulation shows the acquisition of a discrimination between two similar stimuli (i.e., ones sharing elements in common) by two groups. For the control group, both stimuli are novel—that is to say, the weights between the nodes representing the elements of each stimulus and between the stimulus elements and context elements are all set to a baseline level. For the experimental group, the weights between the nodes representing common elements and between those elements and those representing the context are all set to an intermediate level, to represent a moderate level of preexposure to these elements. As can be seen in Figure 9, preexposure to the common elements facilitates acquisition of the discrimination, a result that holds for a wide range of parameters.

Unitization. According to the model, latent inhibition is a consequence of the formation of associations between elements, both between one CS element and another and between contextual and CS elements. But the formation of these associations during the course of preexposure may have other effects on conditioning to a preexposed CS and the extent to which that conditioning generalizes to other, similar stimuli.

Consider again two similar stimuli, AX and BX, presented in a given context, C. Exposure to these stimuli will result in the formation of associations between one A element and another, between A elements and X elements, and between both sets of elements and C elements (and similarly for B and X). After a few trials, the X and C elements will activate the units representing the B elements even on a trial in which AX is presented alone. This can only increase subsequent generalization between AX and BX. Two further mechanisms may, however, serve to counteract this *mediated* generalization. The first is unitization.

Suppose that a rat is inspecting several relatively complex stimuli, such as the black metal triangles and circles used by E. J. Gibson and Walk (1956). One aspect of their complexity is that their features are distributed over space. Consequently, the rat's perception of these features will vary, depending on their point of fixation. Moreover, it is only those features that correspond to local regions of the stimuli (i.e., the details) that will necessarily be sampled in this way. The gross stimulus features—for example, *black*—will be apparent wherever the stim-

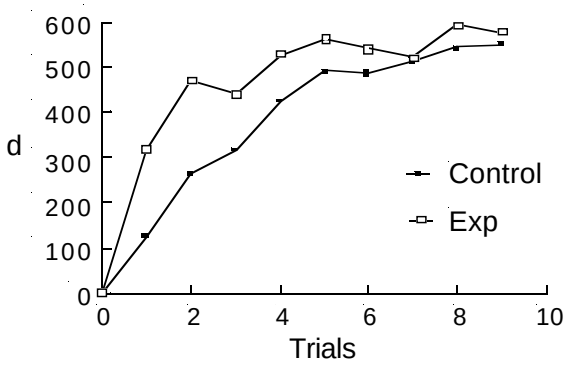

Figure 9. Simulated discrimination learning after latent inhibition to the common, X, elements of a discrimination (Exp) contrasted with learning of the same discrimination when the stimuli are novel. The *d* measure simply indicates the difference in the ability of each conditioned stimulus (CS) to associatively activate the reinforcer node—that is, the activation of the node to CS+ minus the activation of the node to CS−. Learning is more rapid in the Exp group than in the control simulation.

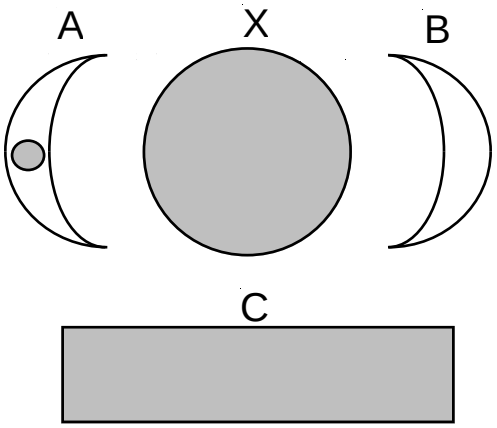

Figure 10. The diagram shows how different sets of elements corresponding to the unique and common elements of two stimuli (as well as the context) are sampled in the simulation whose results are shown in Figure 11.

uli are fixated. If the equilateral triangle and the circle are compared, then at a low resolution (i.e., a coarse level of detail), both will look like black blobs. It is only at a finer level of detail that the discrimination can be made. We suggest that where two stimuli are sufficiently similar and differ only in detail, their unique features will be sampled more variably than their common features. We would not necessarily expect this for a black square versus a white square, since they differ in their *coarse* rather than in their fine detail. But how does the more variable sampling of unique elements lead to a mechanism for perceptual learning?

Figure 10 is an "exploded" version of the earlier Figure 8 and shows A, B, X, and C elements. The shaded region represents the elements sampled and, hence, activated on a given presentation of AX. In line with the above argument, the A elements are shown as being sampled rather variably (i.e., only a relatively small proportion are activated), whereas the X and C elements are shown with a 100% sampling rate. With repeated presentation of stimulus AX, associations form between all the active elements, and as this process is repeated with different samples from A, two important things occur. The X elements will lose salience more rapidly than the A elements, and as a consequence of this, the A elements will tend to develop associations between one another in preference to associations with the X elements. The latter process will enable the currently active subset of A elements to associatively recall other members of that set, and these elements will then be available both for learning and to express any associations that have already accrued to them.

In summary, if the unique elements are variably sampled and the common elements are not, or at least less so than the unique elements, this confers an advantage (in terms of later learning of the discrimination) to the unique elements after preexposure. This advantage could be absolute, rather than relative, if the benefit from unitization outweighs the detrimental effects of latent inhibition. Note that the above argument applies only when the stimuli are "seen" in isolation; if the stimuli were processed simultaneously, the unique elements from both would associate, making it harder to discriminate between them.

Figure 11 shows a simulation of this mechanism in operation. The experimental group has had the nodes representing the unique elements of the stimuli strongly associated to one another; the control group has not. When alternating between the stimuli, the unique nodes are sampled at a 20% rate, whereas the common and contextual nodes are sampled at 100%. Otherwise, the discrimination training was conducted as in the preceding simulation. The figure shows a typical result, rather than an average over many simulations. The experimental group learned faster, and the *d* scores showed great variability from trial to trial, owing to the sampling process.

Formation of inhibitory links. There is a final mechanism for perceptual learning implicit in our model. The formation of excitatory associations between the common and the unique elements of two similar stimuli may well cause mediated generalization between them. But this can be eventually overridden by the formation of some inhibitory associations. If two stimuli, AX and BX, are "inspected" separately, the following contingency is in force. If the A elements are sampled, the B ones are not, and vice versa. Thus, there will be AX trials where A–X associations are formed and BX trials where B–X associations are formed. Moreover, on AX trials, because of the BX trials experienced, there will be an (unfulfilled) expectation of B, and vice versa for BX trials. This negative contingency will result in the A and B elements' forming mutually inhibitory connections. As a result, on an AX trial, the active A elements will suppress any B elements evoked by the contextual or X elements, and on a BX trial, the active B elements will suppress any A elements that might otherwise be evoked. The effect of this

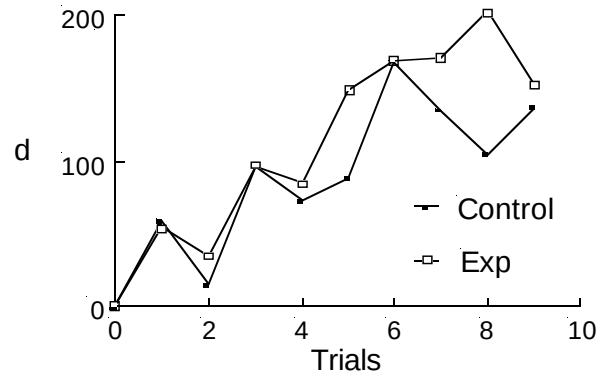

Figure 11. Results of a simulation of unitization. The *d* score once again indicates the difference in activation of the US node to CS+ and CS−. Group Exp (unitized) learned faster than the control group.

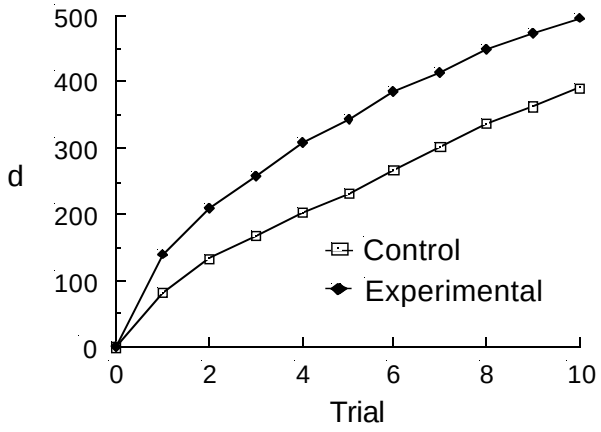

Figure 12. A simulation of the effect of adding inhibitory links between the unique elements of a discrimination. The experimental (inhibitory link) group learned more rapidly than did the controls.

is that if AX is conditioned, there will be little conditioning to the B elements, and on test with BX, the B elements will act to suppress retrieval of the A elements. It might be argued that this formation of inhibitory links will give no advantage over novel stimuli whose elements are not associated to anything. If the stimuli are novel, there will be no B element activation to suppress when AX is presented. However, novel stimuli will begin forming associations between their elements as soon as they are presented and will therefore have to pass through a stage that preexposed stimuli have already passed through (and, in some sense, dealt with). Other things being equal, this would put the novel stimuli at a disadvantage, as compared with the preexposed stimuli.

Figure 12 makes this point explicitly. The only difference between the control and the experimental groups in this simulation is that the experimental nodes representing the unique elements of the stimuli have had the weights between them set to a moderate negative value. This results in a facilitation of discrimination learning.

APPLICATION TO PERCEPTUAL LEARNING: EVIDENCE

Differential Latent Inhibition of Common and Unique Elements

Although E. J. Gibson and Walk's (1956) results were replicated in a number of other, similar studies (e.g., Forgus, 1958a, 1958b), other experiments found no beneficial effect of preexposure on subsequent discrimination learning (e.g., E. J. Gibson, Walk, Pick, & Tighe, 1958; see Hall, 1980, for a review). One factor determining the outcome of these experiments, it soon became apparent, was the difficulty of the discrimination. Preexposure would facilitate the learning of a difficult, but not of an easy, visual discrimination (Oswalt, 1972). The importance of this factor has been confirmed in subsequent studies of perceptual learning in maze discriminations (Chamizo & Mackintosh, 1989; Trobalon, Sansa, Chamizo, & Mackintosh, 1991). For example, Trobalon et al. (1991) preexposed rats to the entire set of extramaze landmarks that defined the spatial location of the arms between which they were subsequently required to discriminate. Such preexposure facilitated the learning of a moderately difficult spatial discrimination—that is, one between two arms separated by an angle of only 45º. But preexposure tended, if anything, to retard the learning of a much easier spatial discrimination between two maze arms separated by an angle of 135º.

Why should the effects of preexposure depend on the difficulty of the discrimination? To answer that question, we need to answer a prior one: What makes a discrimination easy or difficult? Our answer is that two stimuli are easy to discriminate (i.e., there is little generalization between them) to the extent that they share few elements in common; their discriminability decreases as the proportion of common elements increases. Consistent with this analysis, Mackintosh et al. (1991) found that prior exposure to two simple flavors, saline and sucrose, had no effect on the generalization of an aversion from one to the other (i.e., produced no perceptual-learning effect). But similar exposure to two compound flavors, saline–lemon and sucrose–lemon, significantly reduced the generalization of an aversion from one compound to the other.

Why should perceptual-learning effects depend on stimuli sharing common elements? Our first answer, outlined earlier, is that exposure to two or more stimuli sharing common elements will enhance their discriminability by causing differential latent inhibition of their common and unique elements. The argument rests on the seemingly plausible, even incontrovertible, assumption that the magnitude of any latent inhibition effect will be proportional to the amount of exposure to the stimulus or stimulus elements in question (see Elkins, 1973; Fenwick, Mikulka, & Klein, 1975). Because X is exposed twice as often as A or B, it will undergo more latent inhibition. There are, however, grounds for suggesting that other factors might override amount of exposure as a determinant of latent inhibition. According to Pearce and Hall (1980), for example, latent inhibition is a function of the extent to which a stimulus is followed by predictable consequences, and Swan and Pearce (1988) have provided evidence that, with total amount of exposure equated, a stimulus always followed by the same conditioned reinforcer undergoes more latent inhibition than one followed unpredictably by one or the other of two conditioned reinforcers. This suggests the possibility that intermixed exposure to two compound stimuli, AX and BX, might actually generate less latent inhibition to X, which is inconsistently accompanied by either A or B, than latent inhibition to A and B themselves, which are consistently accompanied by X. There has, in fact, been no direct test of this possibility (but see p. 231, where we describe the results of an unpublished study by Trobalon that tend to disconfirm it). When exposure to X was equated, Symonds

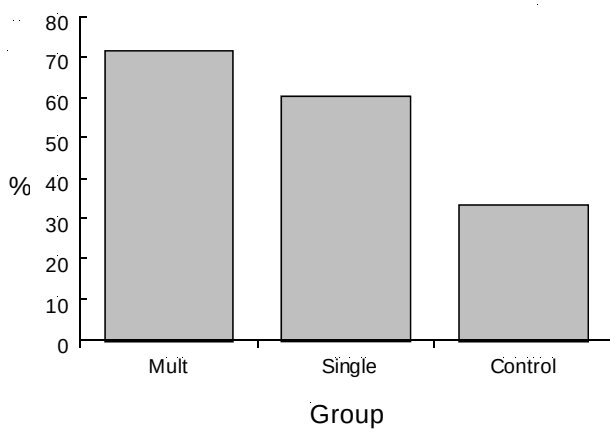

Figure 13. Generalization of an aversion conditioned to one vinegar solution to a second solution (Bennett & Mackintosh, unpublished). Following the preexposure and habituation phases of the experiment, all the animals received two conditioning trials to a solution of either red wine or balsamic vinegar. On the first conditioning trial, all the animals consumed a fixed 8 ml of the conditioned stimulus (CS) solution; on the second trial, they were given 10-min access to the CS. The next day, they received a single 10-min test trial to the solution that had not served as their CS. The data shown are consumption on the test trial, expressed as a percentage of consumption of the CS solution on the second conditioning trial. Statistical analysis confirmed that the control group showed more generalization to the test solution than did either Group Mult or Group Single, which did not differ significantly from one another.

and Hall (1997) found some evidence that latent inhibition was marginally greater when X was consistently accompanied by one flavor than when it was inconsistently accompanied by two, but Bennett and Mackintosh (in press) were unable to confirm even this weak effect.

Preexposure to X alone facilitates discrimination between AX and BX. Be this as it may, a wide variety of experiments have obtained evidence consistent with the basic proposition that, when animals are exposed to two or more stimuli sharing elements in common, differential latent inhibition of their common and unique elements contributes to the perceptual-learning effect observed. The logic of several experiments pointing to this conclusion can be illustrated by a brief description of an unpublished study by Bennett and Mackintosh. In the final phase of this study, three groups of thirsty rats drank a novel vinegar solution, followed by an injection of lithium chloride. The aversion conditioned to this solution was measured on subsequent test trials, as was the extent to which the aversion had generalized to a second novel vinegar solution. The two solutions were red wine and balsamic vinegar, counterbalanced so that, for half of the animals in each group, the aversion was conditioned to red wine vinegar and generalization tested to balsamic vinegar, and for the other half, these assignments were reversed. The three groups were treated differently during the initial exposure phase of the experiment, which lasted for 20 days. Group Mult (for multiple vinegars) drank four different vinegar solutions (garlic, malt, sherry, and cider vinegars), with five exposures to each solution over the 20-day period. Group Single drank a single solution of dilute acetic acid (the flavor common to all the vinegars) on all 20 days. And the control group drank a dilute solution of quinine on each day. These 20 days were followed by 4 days of exposure to the red wine and balsamic vinegar solutions for all animals. This was designed to habituate any neophobia in the control group: On the second trial to each solution, all three groups drank the same amount during the 10-min exposure session. These habituation trials were followed by 2 conditioning days, on which consumption of one of these novel vinegars was followed by a LiCl injection, and a single test trial to the other novel vinegar. The results of this test trial are shown in Figure 13: Since there was a slight, albeit nonsignificant, difference between the three groups on this second conditioning trial, consumption on this generalization test is expressed as a percentage of consumption of the poisoned vinegar on their second conditioning trial. Figure 13 makes it clear that the aversion conditioned to one novel vinegar generalized far more strongly to the other in the control group than in either Group Mult or Group Single (which did not, in fact, differ significantly from one another). The result for Group Mult could be described as showing that familiarization with four different vinegar solutions enhanced rats' ability to discriminate between two new vinegars—a standard perceptual-learning effect. There are, no doubt, a number of theories of perceptual learning that might explain this result, perhaps by appealing to a process of *perceptual differentiation* (J. J. Gibson & E. J. Gibson, 1955) or by suggesting that preexposure drew the animals' attention to the differentiating features of the various vinegars (E. J. Gibson, 1969; although one might wonder why this should enhance the discriminability of two *new* vinegars). But why should constant exposure to the same acetic acid solution in Group Single have proved almost as effective? Latent inhibition provides the obvious explanation. By reducing the salience or associability of the flavor shared in common by all the vinegar solutions, such exposure ensured that when an aversion was conditioned to balsamic vinegar, it was the unique elements of that novel solution that were associated with illness, rather than those it shared in common with the other vinegars (including red wine). And if latent inhibition of the common acetic acid flavor provides the main explanation of the perceptual-learning effect observed in Group Single, it seems plausible to suppose that it provides at least part of the explanation of the similar effect observed in Group Mult.

A number of published studies have provided similar evidence consistent with this analysis. In several other flavor aversion experiments, preexposure to X alone has been as effective as preexposure to AX and BX in reducing generalization between them. By contrast, preexposure to A and B alone does *not* facilitate discrimination between AX and BX (Bennett et al., 1994; Mackintosh et al., 1991). In spatial discrimination studies, prior ex-

posure to the landmarks that defined the spatial location of the various arms of a maze facilitated subsequent discrimination learning only if it involved exposure to the landmarks visible from, and thus common to, the two goal arms. If animals were preexposed *only* to the landmarks unique to each goal arm, this could actually retard subsequent discrimination learning (Rodrigo et al., 1994; Sansa, Chamizo, & Mackintosh, 1996). Thus, in Rodrigo et al.'s (1994) experiment, rats were preexposed either to the landmarks at the end of each maze arm between which they were subsequently required to discriminate or to the landmarks *between* arms and, thus, common to both. Discrimination learning was retarded by preexposure to the landmarks unique to each arm and facilitated by preexposure to the landmarks between the arms. Sansa et al. employed a complementary strategy. Their maze was placed in a circular, black enclosure, with a total of only four distinct landmarks to define the locations of the various arms. For one group, these landmarks were situated at the end of the four arms of the maze; for another, they were placed halfway between each pair of arms. Spatial discrimination learning was retarded in the group preexposed to landmarks situated at the end of each arm but facilitated in the group preexposed to landmarks that lay between arms.

A similar effect has been observed in rats learning to locate the submerged platform in a swimming pool (Prados et al., 1999). In these experiments, the location of the platform, in one quadrant of the pool, was defined by its relationship to four landmarks, placed equidistant around the circumference of the pool. The pool was situated in a circular, black enclosure, and both landmarks and platform were rotated from trial to trial to eliminate any static directional cues. Under these circumstances, when tested with any two or three of the landmarks, rats will still swim directly to the platform, but if only one landmark is left, they swim at random (Rodrigo, Chamizo, McLaren, & Mackintosh, 1997). The implication is that rats use configurations of two or more landmarks to locate the platform. The features shared in common by such configurations of two or more landmarks will, of course, be the features of the individual landmarks themselves, and consistent with this analysis, preexposure to individual landmarks, one at a time, facilitated subsequent learning. But, here again, exposure to pairs of landmarks (i.e., to the relevant configural cues) could actually retard subsequent learning (Prados et al., 1999).

Are variations in the magnitude of latent inhibition correlated with variations in perceptual learning? The results of all these experiments are consistent with the suggestion that exposure to two or more stimuli sharing elements in common enhances their discriminability, and that of other similar stimuli, because it generates more latent inhibition of their common than of their unique elements. In many cases, the basis for this argument is simply that a similar enhancement of discriminability is produced by explicit exposure to the common elements alone. In others, the argument rests on the demonstration that an exposure regime that does not involve exposure to the common elements does not facilitate subsequent discrimination learning. A different line of evidence would involve showing that experimental manipulations that affected the magnitude of latent inhibition had a similar effect on the magnitude of perceptual-learning effects. In practice, however, the argument is not entirely straightforward.

One manipulation that has been thought to have diametrically opposed effects on latent inhibition and perceptual learning is a change of context. As we have seen, latent inhibition is usually disrupted by a change of context between preexposure and conditioning. In the early perceptual-learning studies, on the other hand, such as those of E. J. Gibson and Walk (1956), preexposure took place in home cages, and discrimination training in a different, experimental context. According to Channell and Hall (1981), this feature of their procedure was indeed critical in producing a perceptual-learning effect. When preexposure took place in home cages, Channell and Hall replicated E. J. Gibson and Walk's results; when it took place in the test apparatus, preexposure retarded subsequent discrimination learning. On the face of it, these results contradict our analysis, implying, as they do, that the processes subserving latent inhibition effects cannot be the same as those responsible for perceptual learning.

That implication does not necessarily follow. A preliminary point worth noting is that latent inhibition is not always disrupted by a change of context when the context in which preexposure takes place is as familiar as the home cages employed in E. J. Gibson and Walk's (1956) and Channell and Hall's (1981) experiments (see the experiments by McLaren, Bennett, et al., 1994, described earlier). A further theoretical point is that, in many cases, perceptual-learning effects are doubtless multiply determined. The other processes identified by our theory as contributing to perceptual learning (unitization and inhibitory associations) may not be affected by a change of context in the same way as latent inhibition effects sometimes are. But the most important point to note is that when we are looking at the effects of preexposure to two or more stimuli on their subsequent discriminability, we are saying that it is the *differential* latent inhibition of common and unique elements that contributes to the perceptual-learning effects observed.

Even if a change of context always disrupted latent inhibition, it might still increase the magnitude of any perceptual-learning effect by increasing the *difference* in the amount of latent inhibition accruing to common and unique elements. The argument was spelled out by Trobalon, Chamizo, and Mackintosh (1992) in their discussion of a set of maze experiments, in which unreinforced preexposure to the relevant intramaze cues facilitated subsequent discrimination learning (i.e., yielded a perceptual-learning effect) only when preexposure and discrimination training took place in the same extramaze context. Their analysis is best understood by reference to Figure 14. This illustrates a function relating loss of as-

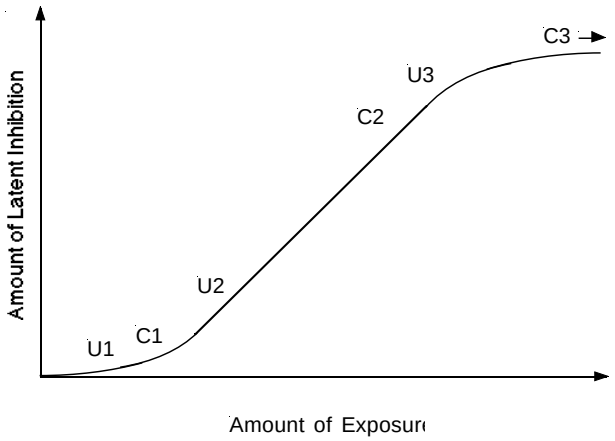

Figure 14. A possible function relating amount of latent inhibition to a stimulus or stimulus element and the amount of exposure. U stands for unique elements, C for common elements of the stimulus. Three exposure durations are shown: 1 = short exposure, 2 = medium exposure, and 3 = prolonged exposure. In each case, the assumption is that the common elements of the stimulus will receive twice as much exposure as the unique elements. Note that the point C3 lies to the right of the graph and, for reasons of scale, cannot be shown in its correct position.

sociability (magnitude of latent inhibition) to the number of preexposure trials. The function shows the loss of associability by common and unique elements of two stimuli following low and high amounts of preexposure, on the assumption that common elements are preexposed twice as often as unique elements. Let us now assume that a change of context disrupts latent inhibition—that is, produces a shift to the left along this function. It is easy to see that after prolonged preexposure, such a shift will tend to increase the difference between common and unique elements: In other words, a change of context will enhance any perceptual-learning effect. After a small amount of preexposure, on the other hand, such a shift might even decrease the differential latent inhibition of common and unique elements. In other words, a change of context under these circumstances may reduce the magnitude of any perceptual-learning effect. As has already been noted, Trobalon et al. did indeed find just such a disruption of perceptual learning by a change of context following the modest amount of preexposure to the discriminative stimuli given in their experiments. They argued that an obvious difference between their procedure and those employed by E. J. Gibson and Walk (1956) and Channell and Hall (1981) was that, in the latter experiments, the total amount of exposure to the discriminative stimuli was at least 50 times as long.

An unpublished study by Trobalon, using the flavor aversion paradigm, provides evidence that supports one of the assumptions underlying this analysis. Trobalon was able to demonstrate not only that latent inhibition is a function of the amount of preexposure given, but also, in line with Figure 14, that the differential latent inhibition of A, B, and X that occurs as a result of preexposure to AX and BX is also a function of the amount of preexposure. After preexposing AX and BX for either 6 or 12 days, an aversion was conditioned to AX, and the animals were tested on BX, X, and A. Reduced generalization to BX, relative to controls, was observed in the preexposed animals. More important, after 6 days of preexposure, there was a strong aversion to A and a weak aversion to X, whereas after 12 days of preexposure, there was little aversion to either. In other words, a modest amount of preexposure to AX and BX generated more latent inhibition to X than to A, but prolonged preexposure produced equally strong latent inhibition to both.

If the analysis illustrated in Figure 14 is correct, it follows that E. J. Gibson and Walk's (1956) and Channell and Hall's (1981) results in no way contradict our account. We must also, of course, accept that, when animals are exposed to two stimuli, AX and BX, it is impossible to predict in advance whether a change of context will disrupt, enhance, or have no effect on the perceptual-learning effect observed. A more determinate prediction about the effect of a change of context on perceptual learning would require a situation in which it was not the *differential* latent inhibition of common and unique elements that contributed to perceptual learning. There is one situation that almost certainly satisfies this requirement. If an aversion is conditioned to one flavor, AX, generalization to a second flavor, BX, is reduced by prior exposure to BX alone (Bennett et al., 1994; Best & Batson, 1977). According to Bennett et al. (1994), a major determinant of this perceptual-learning effect is simply that preexposure to BX results in latent inhibition of X. If that is correct, a change of context between preexposure and conditioning and test phases should both attenuate latent inhibition of X and reduce the magnitude of any perceptual-learning effect (i.e., increase generalization from AX to BX). The results of an unpublished study by Bennett and Tremain confirm this suggestion. The design of their experiment is shown in Table 1, and

Table 1
Design of an Experiment on the Context Specificity of Latent Inhibition (LI) and Perceptual Learning (PL) From Bennett and Tremain (Unpublished)

	Preexposure		Conditioning	Test
Groups	C_1	C_2	(C_1)	(C_1)
LI				
Same	AX	–	AX+	AX
Different	–	AX	AX+	AX
Control	–	–	AX+	AX
	C_1	C_2	C_1	C_1
PL				
Same	BX	–	AX+	BX
Different	–	BX	AX+	BX
Control	–	–	AX+	BX

Note—C_1 and C_2 are two different contexts (different drinking boxes, in different rooms, and sessions run at different times of day). All the animals received four 10-min sessions of preexposure in each context, drinking water where not drinking a flavored solution. Flavors: A, 2% sucrose; B, 0.9% saline; X, 2% lemon juice; +, LiCl injection. All the rats received a single conditioning trial to AX.

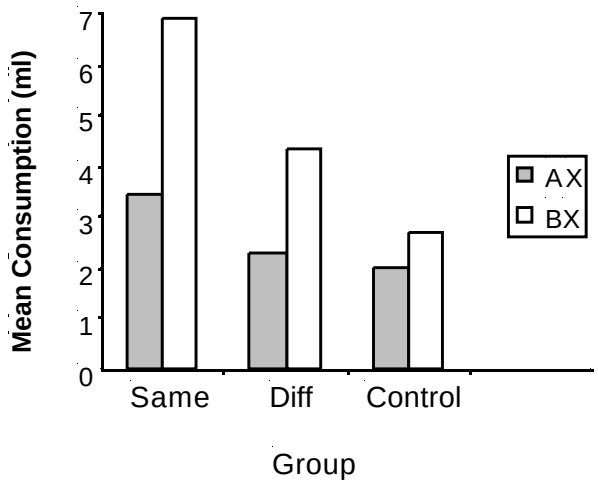

Figure 15. Table 1 gives the design for this experiment (Bennett & Tremain, unpublished). The groups labeled AX were preexposed, conditioned, and tested on AX. Groups labeled BX were preexposed to BX, conditioned to AX, and then tested on BX. A shift in context abolished latent inhibition to AX and also increased generalization from AX to BX (reduced perceptual learning).

their results in Figure 15. The context specificity of latent inhibition was measured in groups preexposed, conditioned, and tested to AX (either in same or in different contexts). The context specificity of perceptual learning was measured in groups preexposed to BX, conditioned to AX, and tested to BX. As can be seen from Figure 15, a change of context between preexposure and conditioning both attenuated latent inhibition to AX and increased generalization from AX to BX. Additional groups, not shown here, established that differences in generalization to BX were perfectly mirrored by differences in the level of conditioning to X alone, thus confirming that it was differences in the degree of latent inhibition to X that determined the magnitude of the perceptual-learning effect.[2]

Unitization

There is ample evidence that latent inhibition of common elements cannot possibly be the only explanation of perceptual-learning effects (e.g., Mackintosh et al., 1991; Symonds & Hall, 1995), and, as was outlined above, we propose two other explanations that arise naturally from the sampling and associative assumptions of the model. The first, unitization, provides one way of capturing the intuitive idea suggested by a number of theorists (e.g., J. J. Gibson & E. J. Gibson, 1955; Hall, 1991; Hebb, 1949) that one effect of exposure to any moderately complex stimulus will be to establish a more detailed and accurate representation of that stimulus.

The process of unitization depends on the assumptions that not all elements of a stimulus will be sampled on a single presentation of that stimulus and that associations will be formed between any simultaneously sampled elements. If the formation of such associations is governed by an error-correcting rule, this will ensure that an initially variable and possibly highly inaccurate representation of a complex stimulus, which includes random, extraneous noise or error, will gradually settle down to become a stable, detailed, and accurate representation. In McClelland and Rumelhart's (1985) words,

> the delta rule can be used to extract the structure from an ensemble of inputs, and throw away random variability. The distributed model acts as a sort of signal averager, finding the central tendency of a set of related patterns. (pp. 167-168)

The impact of this process will be largely confined to relatively complex stimuli since in the case of a simple stimulus, such as a red light or a 1000-Hz tone, we assume that there will be rather little variability in sampling from one presentation of the stimulus to another. In other words, a high proportion of the stimulus's elements will be sampled on each occasion. We assume that sampling is especially selective in those cases in which the various components or aspects of a complex object, scene, or place cannot easily all be apprehended at once. This will be particularly true when they are distributed in space, but it will also apply when different attributes activate different sensory modalities and may also apply when a stimulus with a large number of different attributes (e.g., a complex flavor) is presented for only a brief moment.

Unitization facilitates conditioning. What are the implications of these assumptions for a theory of perceptual learning? Conditioning to a simple stimulus, such as a tone or a light, is retarded by prior exposure to that stimulus—the phenomenon of latent inhibition. But when we are dealing with a more complex stimulus, where variability in sampling becomes a significant factor, the process outlined above may override any loss of associability and mean that prior exposure could actually facilitate subsequent conditioning. A single conditioning trial to a complex stimulus will result in the conditioning of only a subset of its elements, those actually sampled on that trial. If a different, only partially overlapping subset of elements is sampled on the next trial, there will be little evidence of the conditioning that occurred on the first. But some prior exposure to the stimulus will result in the formation of associations between many of its elements. Even if only a subset of elements is sampled on the first conditioning trial and a different subset on the second, these associations will result in the conditioning of unsampled but retrieved elements on the first conditioning trial and the retrieval of some already conditioned but unsampled elements on the second. Although the formation of these associations, and of others between stimulus and contextual elements, will reduce their associability, the latent inhibition effect expected by such a mechanism may be outweighed by this reduction in variability.

There is evidence to support these conjectures. If rats are placed in a novel experimental chamber and given a single brief shock a few seconds later, this one conditioning trial may not be sufficient to establish any evi-

dence of the conditioning of fear to the chamber when the rats are placed back in it the next day for a test trial (Blanchard, Fukunaga, & Blanchard, 1976; Fanselow, 1986; but see Bevins & Ayres, 1995, and Bevins, McPhee, Rauhut, & Ayres, 1997, for evidence that some conditioning can often be detected, especially when a strong shock is used). However, if rats have been exposed to the chamber for a minute or two, without shock, the day before, a single conditioning trial with exactly the same temporal and US parameters will yield substantially stronger evidence of conditioning (Fanselow, 1990; Kiernan & Westbrook, 1993). Thus, prior unreinforced exposure to the chamber, far from retarding conditioning (i.e., generating a latent inhibition effect) can actually facilitate conditioning.

An experimental chamber is surely a complex stimulus, not all of whose features or elements can be sampled simultaneously, and those features sampled on the brief conditioning trial may well differ from those sampled the following day on the test trial. Moreover, if rats are shocked immediately after they are placed in a novel chamber, it will be the transport and handling cues, as much as any features of the chamber itself, that will be associated with shock. This presumably explains why any evidence of conditioned freezing on the test trial is confined to the first few seconds after the rats are placed in the chamber and is not much greater than that displayed by rats who were initially given an immediate shock when placed in a quite different chamber (Bevins & Ayres, 1995).

That unreinforced preexposure to a sufficiently complex CS may actually facilitate, rather than retard, conditioning has been confirmed in a study of flavor aversion conditioning by Bennett, Tremain, and Mackintosh (1996). They found, as one would expect, that prior exposure to a simple flavor, such as a dilute acid or sucrose solution, significantly interfered with the establishment of an aversion to that flavor on a single, brief conditioning trial. But when they used a more complex flavor as the CS—a mixture of monosodium glutamate, sucrose, and quinine—and gave only a single, brief conditioning trial, prior unreinforced exposure to the CS significantly facilitated conditioning.

Unitization can reduce generalization. We have hitherto defined perceptual learning as an enhancement of discrimination, or reduction in generalization, between two or more stimuli as a result of prior exposure to those or similar stimuli. The evidence cited so far in this section has simply been concerned with an increase in the rate of conditioning to a complex stimulus as a result of prior exposure. Is there any reason to expect that the process of unitization will reduce generalization between two or more stimuli? Kiernan and Westbrook (1993) did in fact find just such an effect. Prior exposure to a chamber, in which rats later received a single conditioning trial, not only increased the level of conditioned fear or freezing observed when rats were placed back in the conditioning chamber, but also reduced the level of freezing observed when the rats were tested in a somewhat different chamber. The enhancement of conditioning to the CS was accompanied by a reduction in generalization to another stimulus (recall that Bevins & Ayres, 1995, also observed substantial generalization from one chamber to another in rats given no preexposure).

Why should unitization lead to any decrease in generalization from one complex stimulus to another? As we have already argued, the answer depends on the assumption that the process of sampling is not a random one. If the elements of a complex stimulus sampled on any trial were simply a random subset of the whole, unitization might speed up conditioning but could have no effect on generalization: Any *randomly* sampled subset of elements of the CS would share exactly the same proportion of elements in common with a second stimulus as the complete set did. Unitization will reduce generalization only if the initial sampling of a complex CS is biased toward those elements it shares in common with the stimulus to which generalization is being measured. Any such bias will mean that the conditioning that occurs on the first trial to the novel CS will generalize strongly. But the process of unitization, by establishing associations between all the elements of the CS, will ensure that more of the unique elements will be retrieved, even if they are less likely to be sampled, and therefore, that generalization to other stimuli will be reduced.

The assumption that initial sampling may be biased toward elements that one complex stimulus shares in common with others seems, in many cases, a plausible one. The chamber in which rats were shocked in Kiernan and Westbrook's (1993) experiment was roughly the same overall shape (although differing slightly in size) and had exactly the same grid floor as the chamber in which they were tested for generalization. It seems likely that the overall shape would be one of the first aspects that rats would notice and that the grid floor would be the feature most strongly associated with shock. The other differences (exact dimensions, the material from which the walls were made) were surely less salient. By the same token, when rats are asked to discriminate between visual patterns, such as a circle and a triangle (which, no doubt, seem simple to us), it seems reasonable to suggest that the detection of those features that differentiate them, such as the corners of the triangle or the curvature of the circle, will require careful inspection, whereas a casual glance will be more likely to sample features (color, texture, solidity) of the two that they may share in common. We acknowledge that our analysis here rests on little more than considerations of plausibility. What is needed in order to test its validity, of course, is the construction of stimuli where such assumptions are less plausible or where exposure somehow biases animals toward sampling some features rather than others.

Search images and weight decay. Our model assumes that the weight changes that occur as a result of the simultaneous activation of two or more units contain both a transient and a permanent component (see above). Among other things, this allowed us to explain some of

the effects of distribution of practice (see our discussion and simulation of massed vs. spaced preexposure to a stimulus, p. 220). One such effect is observable in research on search images and is readily explained as a consequence of the transient component of the weight changes underlying the process of unitization. The concept of a search image has been used to explain the well-documented observation that birds searching for cryptic prey overselect more abundant prey. If two types of food, A and B, are distributed at random across a foraging site, in the ratio 75 As to 25 Bs, the first 1OO items of food collected will typically consist of significantly more than 75 As and significantly fewer than 25 Bs (e.g., Tinbergen, 1960). There is good evidence that this reflects an increase in the detectability of the more abundant food: If birds are trained on visual discrimination problems, being rewarded for pecking on trials when one or other of two camouflaged targets, A or B, is presented on a screen, but not in the absence of a target, they respond more accurately to the more frequently occurring target (see, e.g., Pietrewicz & Kamil, 1979; Plaisted & Mackintosh, 1995). The popular explanation of these results is that the predator forms a search image of the more abundant prey or more frequently occurring target, which simultaneously enhances the detectability of that item and interferes with the detection of the less abundant prey or less frequent target (Tinbergen, 1960).

We believe that the process of unitization provides a sufficient account of such results and that the concept of a search image may well contain surplus baggage. Unitization is sufficient to explain why a cryptic target becomes easier to pick out from its background with sufficient practice because it results in the establishment of a complete and accurate representation of the target, which will be retrievable even when only a few of its elements are activated.[3] Moreover, the transient component of the weight changes underlying unitization will explain the transient nature of search image effects. Following a session in which A and B occur in the ratio 75 As to 25 Bs and birds respond more accurately to As than to Bs, they will soon start responding more accurately to Bs than to As in an immediately succeeding session in which this ratio is reversed to 25 As to 75Bs (Plaisted & Mackintosh, 1995). As the frequency of A trials decreases, so the weight changes underlying unitization of A begin to decay; conversely, the increase in the frequency of B trials now ensures a transient increase in unitization of B.

There is nothing in the concept of unitization, however, corresponding to the idea implicit in the search image concept, that a search image for A *prevents* the bird from detecting B. Interestingly enough, it may be quite unnecessary to appeal to any such interference mechanism. Plaisted (1997) showed that search image effects may be attributable to the fact that, other things being equal, the average interval of time elapsing between two successive encounters with a high-density prey is bound to be shorter than that elapsing between two encounters with a low-density prey. When Plaisted equated the interval between two consecutive trials with a high-frequency target and that elapsing between two trials with a low-frequency target, the entire search image effect disappeared: High- and low-frequency targets were detected with the same level of accuracy. The sole determinant of discriminative performance was the time that had elapsed since the last presentation of that stimulus. Her results seem entirely consistent with our account of unitization. Since the targets were hard to detect, being small black-and-white patterns appearing on a black-and-white checkerboard background, their detection should have benefited from the establishment of an accurate representation. Each encounter with a target should strengthen the connections between units activated by its various features, but if these weight changes decay over time, some of the benefit from an encounter will be lost as the intertrial interval (ITI) to the next trial increases. Of course, as Plaisted and Mackintosh (1995) also found, with prolonged practice with a particular pair of targets, discriminative performance, even at very long ITIs, slowly improved. But that is predicted from the assumption that part of any weight change underlying unitization is permanent.

Inhibitory Associations

Other things being equal, it is obvious enough that the formation of associations between the elements of a stimulus (i.e., the process of unitization) might just as easily increase as decrease generalization from that stimulus to another. If two stimuli, AX and BX, contain elements in common (X), the formation of associations between A and X elements and between B and X elements could have either or both of the following consequences. On a conditioning trial to AX, X would retrieve B elements that might then be associated with the US. Equally, on a test of generalization to BX, X would retrieve the already conditioned A elements, and the magnitude of any CR would thereby be augmented. Each of these effects would serve to increase generalization from AX to BX. A final step in our associative analysis shows how such mediated conditioning and generalization effects may, after sufficiently prolonged preexposure to AX and BX, be counteracted by the development of mutually inhibitory associations between the two sets of unique elements, A and B.

Presentations of AX and BX during the preexposure phase of a perceptual learning experiment will surely result in the formation of excitatory associations between A and X and between B and X. But if the schedule of preexposure intersperses or alternates presentations of the two stimuli, as in typical studies of perceptual learning in flavor aversion, the negative correlation between the presence of A elements and that of B elements will mean that, on a trial in which A units are externally activated (by the presence of A), B units will be internally activated (by input from X), but not externally activated. This is the condition that will allow the formation of inhibitory connections from A to B (and vice versa from B

to A, on BX trials). In language made more familiar by the Rescorla–Wagner (1972) model, the presence of A signals the absence of the otherwise predicted B, and this negative discrepancy between obtained and predicted outcome is what generates inhibitory conditioning. With sufficient exposure to AX and BX, these inhibitory connections should counteract the mediated conditioning or generalization generated by the excitatory connections between A and X and between B and X. Thus, on a conditioning trial to AX, although X will activate B, A will suppress this activation. And on a test trial to BX, although X will activate A, B will suppress this activation.

Perceptual learning is enhanced by conditions favoring the formation of inhibitory associations. Although the formation of such mutually inhibitory associations seems a straightforward consequence of the application of most modern associative theories, including Rescorla and Wagner (1972), Wagner (1981), and our own model, we acknowledge that the evidence that they play an important role in perceptual-learning effects remains largely indirect. But it is not inconsiderable. First, a variety of experiments have shown that perceptual learning is enhanced by conditions that will, according to such theories, generate strong inhibitory associations between the unique elements of two compound stimuli. Inhibitory conditioning depends on the absence of an otherwise predicted event. Separate presentations of AX and BX result in mutually inhibitory associations between A and B because of the prior formation of excitatory associations between X and A and between X and B. Perceptual learning should therefore depend on two stimuli's sharing elements in common. One reason for this, as we have seen, is that prior exposure to the two stimuli will result in differential latent inhibition of their common and unique elements. But if discrimination between AX and BX is facilitated both by latent inhibition of X and by the establishment of mutually inhibitory associations between A and B, it will benefit not only from prior exposure to AX and BX, but also from prior exposure to AY and BY—that is, to two other compound stimuli sharing a quite different set of elements in common. Mackintosh et al. (1991, Experiment 4) confirmed this prediction in a study of flavor aversion conditioning. Experimental groups were exposed either to AX and BX or to AY and BY, whereas control groups were exposed to AX and BY or to AY and BX (where A and B were sucrose and saline, and X and Y lemon and quinine). Each group also received sufficient exposure to X or Y alone to ensure that all the groups received the same total amount of exposure to these two flavors. In the final stage of the experiment, the two experimental groups learned the discrimination between AX and BX at the same rate, and significantly faster than the control groups who had been exposed to A and B in the absence of any common flavor.

The development of mutually inhibitory associations between A and B must also depend on the precise schedule of exposure to AX and BX. Following an earlier study of imprinting by Honey, Bateson, and Horn (1994), which showed a similar effect, Symonds and Hall (1995, 1997) found that alternating or intermixed exposure to two flavors, AX and BX, was more effective in reducing generalization from one to the other than was a *blocked* schedule in which animals received exactly the same total amount of exposure to AX and BX but in which all trials with AX preceded those with BX (or vice versa). Interspersed trials with AX and BX would be expected to establish mutually inhibitory associations between A and B. But if all the AX trials precede BX trials, A will not be established as an inhibitor of B (since B is not yet predicted by the presence of X), and although, in principle, inhibitory associations from B to the now absent A might be formed on BX trials, evidence from other types of experiment suggest that they would be, at best, very weak (e.g., Ellis, 1970).

The Espinet effect. A quite different line of evidence has suggested that after intermixed exposure to AX and BX, not only is there a reduction in generalization from one compound to the other, but also, if A alone is then paired with a US, B will now act as a conditioned inhibitor of that US. This rather striking finding was first reported in flavor aversion conditioning by Espinet et al. (1995): After preexposure to AX and BX, they conditioned an aversion to flavor A alone; when flavor B was subsequently paired with the lithium US, the rats were slow to acquire an aversion to it (a retardation test of conditioned inhibition); if another flavor, C, was paired with lithium, the aversion conditioned to C was alleviated by adding B to C (a summation test). The control groups against which these inhibitory effects were assessed included (1) different preexposure regimes—either exposure to A and B alone without a common X element or a much smaller amount of exposure to AX and BX (recall that inhibitory associations between A and B will only form after the establishment of excitatory associations between X and A and between X and B)—and (2) either unpaired presentations of A and the lithium US or pairing A with a saline injection. Leonard and Hall (in press) have confirmed Espinet et al.'s basic findings, using a different conditioning procedure (conditioned suppression), and have further shown that the effect depends on intermixed rather than blocked preexposure to AX and BX (a finding also reported by Bennett et al., 1999; see p. 236).

The controls employed in these experiments all suggest that their results depend on the establishment of inhibitory associations between A and B during the course of preexposure. But why should such mutual inhibition between A and B turn B into a conditioned inhibitor of a US subsequently paired with A? Although Espinet et al. (1995) suggested a different possibility, perhaps the simplest explanation would run as follows: If B inhibits A, not only will it suppress the activation of A, it will also suppress the activation of any associate of A—in this case, the US. The reasoning here is that units representing B, by virtue of their inhibitory links to those representing A, will pass negative input to units for A, sending their

activation negative. The excitatory links from A units to US units, coupled with negative activation of the units representing A, will now tend to counteract any excitation of the US units. On a summation test with C and B, although C will activate the US representation, B will prevent this. On a retardation test, the excitatory associations between B and the US that tend to activate the US will be counteracted by this inhibitory effect. Finally, we can return to the effect of preexposure to AX and BX on the subsequent generalization to BX of an aversion conditioned to AX. If the AX conditioning trial causes B to suppress activation not only of A, but also of the US paired with AX, there will, of course, be little or no generalized aversion to BX.

Although this is all necessarily somewhat speculative, Bennett et al. (1999) have provided additional evidence for a crucial step in the argument. According to the theory, intermixed exposure to AX and BX should establish mutually inhibitory associations between A and B. But which of the two inhibitory connections is the more important in reducing generalization from AX to BX—that from A to B or that from B to A? According to the above analysis, such a reduction in generalization depends on the fact that after conditioning to AX, B will suppress activation not only of A, but also of the US associated with A. This implies that it is the inhibitory connection from B to A that is important. Bennett et al. (1999) tested this argument by devising preexposure schedules that should, in principle, have established unidirectional inhibitory connections. In addition to groups given standard intermixed or blocked preexposure to AX and BX, two additional groups received one presentation of each compound in each daily preexposure session. Group AX→BX always received AX first, followed a minute or two later by BX, whereas Group BX→AX always received the two solutions in the opposite order. The backward pairing of BX with AX in Group AX→BX should have established an inhibitory association from B to A, since B now signals the absence of A until the following preexposure trial, a minimum of 4 h later. Conversely, the delayed forward pairing of AX with BX experienced by this group might have interfered with the establishment of any inhibitory association from A to B. The parallel argument implies that, in Group BX→AX, a strong inhibitory association only from A to B would be established. In agreement with the argument that it is the inhibitory association from B to A that reduces generalization from AX to BX, Group AX→BX showed as little generalized aversion to BX as the group given intermixed preexposure to the two solutions, whereas Group BX→AX showed as much generalization as the group given blocked preexposure. In an additional set of experiments, Bennett et al. showed, both with retardation and summation tests, that it was preexposure to AX followed by BX, rather than the other way around, that turned B into a conditioned inhibitor of the US paired with A.

Direct evidence of inhibition between A and B following exposure to AX and BX. The evidence for the role of inhibitory associations between A and B in reducing generalization from AX to BX has been largely indirect and circumstantial. Is there any direct evidence to establish that these inhibitory associations are formed? What would such evidence look like? Consider the case in which A is a saline solution and B a sucrose solution. It is well established that sodium depletion (i.e., the induction of a need for salt) will increase a rat's consumption not only of a saline solution, but also of another solution previously associated with saline (e.g., Fudim, 1978; Rescorla & Durlach, 1981). Thus, if we gave rats a paired presentation of saline and sucrose, we should expect the induction of a sodium appetite to increase subsequent consumption of the sucrose solution as well. If this increase in consumption depends on the formation of an excitatory association between sucrose and saline, prior establishment of an inhibitory association between the two should attenuate the effect of sodium depletion on sucrose consumption. In effect, this would constitute a retardation test of inhibitory conditioning. In unpublished experiments Mackintosh and Bennett gave rats either intermixed or blocked preexposure to saline–lemon and sucrose–lemon solutions, followed by a single pairing of saline and sucrose. Subsequent sodium depletion increased consumption of sucrose more in the blocked group than in the alternating, intermixed group. Control groups established that this difference between the two groups depended on their having experienced a saline–sucrose pairing and that it disappeared after a sufficient number of pairings (prior inhibitory conditioning merely retards subsequent excitatory conditioning: it does not permanently prevent it).

PERCEPTUAL LEARNING: OTHER ACCOUNTS

The phenomenon of perceptual learning should have been familiar to psychologists since the days of William James, but theoretical analysis has lagged far behind everyday observation and laboratory demonstration. The account outlined above represents one of the few attempts to provide a systematic, reasonably well-specified, and experimentally testable analysis that rests on well-documented empirical foundations. Its ability to account for the phenomena of perceptual learning is one of the main achievements of our model. A brief review of possible alternative accounts will, we believe, reinforce that claim.

Associative Theories

We are, of course, far from being the first to advance an associative account of perceptual learning. Empiricist theories of perception, which assume that past associative learning somehow bridges the gap between raw sensation and finished percept, have a long and venerable history in philosophical and psychological thought. To take a somewhat more recent example, the process that we have referred to as unitization bears more than a passing resemblance to Hebb's (1949) concept of the cell assembly. According to Hebb, the perception of any object, even

one as simple as the drawing of a triangle, depends on the elaboration of associative links between units responding to particular features. Fixation on one corner activates one set of units; when the eyes move to fixate a second corner, the new units now activated become associated with those activated by the first. With sufficient experience, inspection of any one part of the drawing activates the entire set of units, the cell assembly, representing the triangle.

William James (1890) proposed a quite different associative account of perceptual learning. The example he used to illustrate his theory was that of a person learning to discriminate between claret and burgundy. Although initially the flavors of the two classes of wine are barely distinguishable, they each become associated with different labels and contexts (the name *claret* with that wine that I drank on such and such an occasion), and the resulting discrimination between A (flavor of claret) with C (labels and contexts associated with claret) and A′ (flavor of burgundy) with C′ (labels and contexts associated with burgundy) is very much easier than that between A and A′ alone.

Hall (1991) has reviewed evidence consistent with what has since come to be known as acquired distinctiveness theory and with its obverse, the theory of acquired equivalence and mediated generalization. There can be little doubt that the effects predicted by these theories often occur, although, as Hall notes, it is not always easy to predict which theoretical mechanism will be the more influential in a given experiment. The reality of acquired equivalence and distinctiveness effects, however, does not imply that they can explain how unreinforced exposure to two similar stimuli should enhance their discriminability. The circles and triangles hanging from the walls of the rats' home cages in E. J. Gibson and Walk's (1956) experiment were both presented, without further consequence, in the same context (the exact position of the stimuli was varied from day to day). When thirsty rats are preexposed to two flavored solutions, the two solutions are again presented in the same context and are, presumably, associated with the same reinforcing consequence—namely, a reduction in thirst. The principle of acquired equivalence implies that preexposure should increase generalization between these stimuli, not enhance their discriminability. Indeed, there is direct evidence that the discrimination between the two compound flavors, AX and BX, is facilitated more by preexposure to A and B in compound with another common feature, Y (i.e., to AY and BY), than by preexposure to A and B each paired with a distinctively different flavor (i.e., AX and BY, or vice versa; Mackintosh et al., 1991). But it is this second group that should benefit from any acquired distinctiveness effect, whereas the first should experience an acquired equivalence effect.

To drive the point home, it is possible to demonstrate perceptual-learning effects in a situation that explicitly *requires* subjects to associate a set of stimuli with a common consequence (Aitken, Bennett, McLaren, & Mackintosh, 1996: McLaren, Leevers, & Mackintosh, 1994). In these experiments, the preexposure phase involved subjects' learning a categorization task. Variable exemplars of two categories were generated by random distortions of the central prototypes of each category (the prototypes were two black-and-white checkerboard patterns, and the exemplars were generated by changing a randomly chosen subset of the squares from black to white or vice versa). In the categorization phase, the participants were shown one exemplar at a time and learned, with feedback, to assign it to its correct category. This training explicitly required subjects to associate all exemplars of one category with a common consequence. Thus, acquired equivalence theory predicts that the experience of categorization should make it harder, rather than easier, to learn a new discrimination between two new exemplars of one of these categories. In fact, both people (McLaren, Leevers, & Mackintosh, 1994) and pigeons (Aitken et al., 1996) found it easier to discriminate two new instances of a familiar category than one they had never seen before (although in pigeons, it was also possible to discern evidence of an acquired equivalence effect working against this outcome).

Our own explanation of this finding sees it as another case of differential latent inhibition of common and unique elements. In order to learn the categorization problem, subjects must, it is true, associate those features or elements diagnostic of category membership with the appropriate category. These elements are, of course, those common to the prototype and to all exemplars of the category. But because most of these elements appear on every trial, at the same time as they are becoming associated with the appropriate category they are also losing salience. By the end of categorization training, they will be strongly associated with the correct category—but will be slow to enter into any new association. By contrast, the elements unique to each exemplar will have received relatively little exposure, and their salience will remain high. Thus when subjects are asked to discriminate between two new exemplars of one category, even though they will be similar to exemplars they have previously categorized together, they will also be easy to tell apart, since the salience of the features they share in common has sharply decreased, whereas that of the features that differentiate them remains high.

Two additional pieces of evidence are consistent with this analysis (which is, of course, exactly the same as that provided for the vinegar categorization experiment described earlier, p. 229). First, Aitken et al. (1996) showed that discrimination between two exemplars of a category was similarly enhanced by prior discrimination training between the prototype of the category and a second checkerboard pattern. Second, McLaren (1997) and A. J. Wills and McLaren (1998) have shown that the beneficial effect of categorization on subsequent discrimination between new exemplars of a familiar category critically depends on the structure of the category. McLaren (1997) studied the effect of familiarizing subjects with

two different categories of checkerboards. In one case, the category was defined by a prototype, with exemplars generated by adding noise to the prototype (as in the Aitken et al., 1996, and McLaren, Leevers, & Mackintosh, 1994, experiments just discussed). In the other case, he started with a master checkerboard and generated exemplars by randomly shuffling or permuting rows of the master pattern. In the latter case, there was no prototype, defined as the average or central tendency of the exemplars, because the average of the exemplars generated by this method consisted of a number of columns of different shades of gray and so was not itself a member of the category, which contained only checkerboards. The effects of familiarization with these different category structures were quite different (even though both types were equally easy to train as categorization problems). As before, the prototype-defined categories showed a strong perceptual-learning effect, whereas the shuffled categories did not. The results of A. J. Wills and McLaren (1998) were equally compelling; using a free-classification paradigm, they were able to show perceptual learning contingent on preexposure for prototype-defined categories, but the reverse effect for the categories generated by randomly permuting rows of some master pattern.

The importance of these results is that they directly contradict any suggestion that preexposure must inevitably result in perceptual learning, given the procedures used in these experiments (as would, perhaps, be expected on the basis of some theories of perceptual learning). On the contrary, these procedures only give rise to perceptual learning with a particular type of category structure—namely, a prototype-defined category. It is this structure that ensures that differential latent inhibition of common and unique elements can occur, with the prototypical features shared by exemplars of a given category playing the role of common elements and the *noise* features that define an exemplar playing the role of unique elements.

Differentiation Theory

J. J. Gibson and E. J. Gibson (1955) advanced a rather different set of objections to acquired distinctiveness theory. Perceptual learning, they argued, must surely be a matter of establishing more accurate, veridical representations of stimuli, but acquired distinctiveness theory implies just the opposite. If perceptual learning is a matter of associating the representations of two stimuli with different labels and contexts, this implies that "perception is progressively in decreasing correspondence with stimulation." According to their differentiation theory,

> We learn to perceive in this sense: that percepts change over time by progressive elaboration of qualities, features and dimensions of variation. . . . Perceptual learning, then, consists of responding to variables of physical stimulation not previously responded to. . . . The observer sees and hears more . . . because he discriminates more. He is more sensitive to the variables of the stimulus array. (p. 34)

The idea seems plausible enough. When we first drank red wine, we probably noticed little more than that it was somewhat unpleasant, with a faint taste of black currant and grape juice, but notably less sweet. The connoisseur detects hints of other fruits beside black currant, will tell you whether it is full or medium bodied, soft or with a tannin *bite*, and so on. Thirsty rats, given a novel, complexly flavored solution to drink, may at first notice only that it is funny tasting water, which has the desirable property of quenching their thirst. With sufficient experience, however, they come to detect its more subtle properties and thus to discriminate it from the other flavored solution they are given to drink on alternate days.

All this may be true. But it hardly constitutes an *explanation* of perceptual learning, for it does not address the question of why we (and rats) should eventually come to notice the subtle features of a stimulus that initially escaped our attention. What is the mechanism that drives this learning? J. J. Gibson and E. J. Gibson (1955) provided no hints. But in later writings E. J. Gibson (1969; E. J. Gibson & Levin, 1975) suggested that perceptual differentiation involved noticing the distinctive and contrasting features that served to differentiate one stimulus from another and that the learning process involved was one of learning to abstract and attend to these features and to ignore other irrelevant features that failed to distinguish one stimulus from another.

Theories of attention in discrimination learning have long argued that people and other animals will solve a discrimination problem by learning to attend to relevant and to ignore irrelevant stimuli, cues, and dimensions (e.g., Lovejoy, 1968; Mackintosh, 1975; Sutherland & Mackintosh, 1971; Zeaman & House, 1963). But all such formal theories have assumed that differential reinforcement is necessary, in one way or another, to drive attentional learning. No such differential reinforcement is provided by the experimenter when rats are preexposed to circles and triangles hanging from the walls of their cage or given different flavored solutions to drink on alternating days. Indeed, as E. J. Gibson and Levin (1975) acknowledged, the experimenter can not possibly provide such differential reinforcement, since he has no way of knowing when the subject is attending to relevant, differentiating features of the stimuli or to their irrelevant, common features. They argued that the reinforcement must be intrinsic, a matter of a reduction in the subject's uncertainty or of an increase in his control over the environment. Such a notion may have some validity when applied to human subjects in psychological experiments. But to apply them to rats in the kind of preexposure experiment we are concerned to understand comes perilously close to mere hand waving.

It does, however, seem possible to bring experimental evidence to bear on the issue. The categorization experiments described in the context of acquired distinctiveness theory (p. 237) seem equally to raise problems for a theory of perceptual learning that appeals to an active process of attention. In these experiments, subjects were ini-

tially required to discriminate the variable exemplars of one category from those of another, where each exemplar was generated by random distortion of the prototype defining its category (Aitken et al., 1996; McLaren, 1997). For the purposes of this categorization task, the relevant features of each exemplar are those it shares in common with its prototype and other exemplars belonging to that category. The irrelevant features are those unique to each exemplar. Thus, attentional theories must predict that subjects will learn to attend to the features common to all the exemplars of each category and to ignore their unique features. But training on this categorization problem made it easier, not harder, for subjects to subsequently discriminate between two new exemplars of one of the categories, even though they should have learned to attend to those features that were irrelevant for this new discrimination.

As we have already noted, our own explanation of these results appeals to the concept of latent inhibition. Features common to all the exemplars of a category will receive more exposure during the course of categorization learning than will those unique to each. Although relevant to the solution of the categorization problem, they will undergo more latent inhibition, thereby rendering them relatively less salient by the time subjects are required to discriminate between exemplars drawn from the same category. The point about latent inhibition as an explanation of certain perceptual-learning effects is that, unlike changes in attention, it does not require differential reinforcement for its development. It is an automatic consequence of mere exposure to a stimulus.

E. J. Gibson (1969) has, however, argued that some changes in attention may not require differential reinforcement. She suggested that the simple opportunity to compare and contrast two similar stimuli would automatically direct attention to their differentiating features and away from those they shared in common. Her suggestion has been taken up by Honey et al. (1994) and Symonds and Hall (1995) as an alternative explanation of the finding that alternating or intermixed exposure to two stimuli, AX and BX, resulted in less generalization from one to the other than did blocked exposure with all AX trials preceding BX trials (or vice versa). It seems reasonable to suppose that alternation back and forth between the two stimuli would be necessary to engage E. J. Gibson's comparison process. Indeed, what E. J. Gibson seems to have had in mind was the case in which subjects were exposed simultaneously to two stimuli, side by side, and could scan back and forth, noting the contrast between them. In Symonds and Hall's (1995) flavor preexposure experiment, on the other hand, the minimum interval separating presentation of one flavor from that of another was 4 h. An appeal to E. J. Gibson's process seems rather less plausible here. At the very least, it must surely follow that a shorter interval between presentation of one flavor and that of another would be more likely to engage a process of comparison and hence to reduce generalization between the two. Bennett and Mackintosh (in press), however, found absolutely no evidence of any such effect. In agreement with Symonds and Hall (1995), they observed less generalization from AX to BX following intermixed than following blocked preexposure, but variation in the interval separating presentation of the two flavors from 2 min to several hours had no effect on generalization. Moreover, this difference between alternating and blocked preexposure in generalization from AX to BX was not accompanied by any difference in the level of conditioning to the common element, X. E. J. Gibson's analysis implies that the beneficial effect of alternating exposure arises because the opportunity for comparison it provides will direct attention toward A and B, and according to most formal theories of attention, any increase in attention to A and B will be at the expense of attention to X. Thus, it seems to predict that any reduction in generalization will be accompanied by a decrease in the level of conditioning to X.

Although Bennett and Mackintosh's (in press) findings are surely inconsistent with E. J. Gibson's (1969) analysis, that analysis does seem to rest on a reasonable enough assumption—namely, that the act of scanning back and forth between two similar stimuli somehow filters out the features they share in common and, by some contrast process, renders their distinctive, unique features still more distinctive. We accept that some such process may well operate but suggest that it may have nothing to do with the sort of perceptual-learning effects we are concerned with here. A variety of experiments have shown that if animals are required to learn a visual discrimination between two similar stimuli, they find it easier if the two stimuli are presented simultaneously, side by side, on the same trial than if they are presented separately, one at a time, on successive trials (e.g., McCaslin, 1954; Saldanha & Bitterman, 1951; S. J. Wills & Mackintosh, 1999). Such results have often been taken as evidence of relational learning. Simultaneous presentation of two stimuli differing, say, in brightness generates a relational cue (one is brighter, the other darker) that can be used to facilitate their discrimination. But a simpler explanation appeals to nothing more than a lower level process of sensory adaptation and contrast (Riley, 1958; S. J. Wills & Mackintosh, 1999).

Consider two similar stimuli (similar because they share common elements), AX and BX, which are presented simultaneously, so that subjects scan back and forth between them or successively in very rapid alternation: AX–BX–AX–BX, and so forth. It follows that the interval between each effective presentation of A or B must be as long as it takes the experimenter to present, or the subject to inspect, the other stimulus. But the interval between each presentation of X can be effectively zero. A transitory process of sensory adaptation will reduce the effective salience of X more than that of A or B, thus rendering the two stimuli more discriminable. But this increase in discriminability will last only while the two stimuli are being presented in this manner. Consistent with this, S. J. Wills and Mackintosh (1999) found

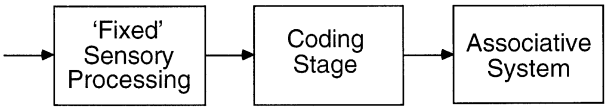

Figure 16. A simple generic model of perceptual learning employing a coding stage.

that pigeons learned a difficult brightness discrimination more rapidly with simultaneous than with successive presentation of the two stimuli but that the performance of birds trained with simultaneous presentation was immediately disrupted when the stimuli were presented successively. If such adaptation or contrast effects operate only at the time at which two stimuli are benefiting from the opportunity for simultaneous comparison, they are unlikely to explain long-term perceptual-learning effects. Indeed, they may be inimical to any such long-term effect.

Transient sensory adaptation is represented in a theory such as SOP (Wagner, 1981) by the notion of short-term habituation or self-generated priming. And according to SOP, short-term habituation can disrupt long-term habituation (e.g., Davis, 1970). Our own model, although for slightly different reasons, makes exactly the same prediction about the effects of trial spacing on latent inhibition: Highly massed preexposure, although generating greater short-term latent inhibition, will actually disrupt long-term latent inhibition. Our assumption that weight changes contain a transitory as well as a permanent component implies that if a second learning episode occurs before the transient weight changes from a first episode have had time to decay, the magnitude of any weight changes on the second episode will be attenuated (this follows from the application of an error-correcting learning rule). Since latent inhibition in our model is associative, dependent on the establishment of internal inputs to the units representing a stimulus, it follows that highly massed presentations of a stimulus during preexposure will disrupt long-term latent inhibition to that stimulus.

A corollary of this analysis is that *very* rapid alternation between two stimuli, AX and BX, during preexposure might actually disrupt perceptual learning (i.e., increase subsequent generalization from AX to BX). This is exactly what Bennett and Mackintosh (in press) observed. Although a reduction in the interval separating presentations of AX and BX during preexposure from several hours to 2 min had no effect on subsequent generalization between the two, a further reduction to a nominal 0 sec significantly *increased* generalization. Moreover, it did so by increasing the strength of conditioning to the common X element. Studies of imprinting (Honey & Bateson, 1996; Honey et al., 1994) have also reported similar effects on perceptual learning.

Connectionist Versions of Differentiation Theory

The Gibsonian position implies that there is some perceptual-learning process that leads to progressive differentiation of stimuli with experience. We have criticized this suggestion for being vague and imprecise, but it can, in fact, be implemented by inserting, prior to any associative system, a coding stage that develops appropriate codes for stimuli as a result of exposure to them. Figure 16 depicts this schematically. A given stimulus input is processed by a variety of hardwired, or *fixed*, sensory mechanisms to provide primitive information for the coding stage. This stage develops a set of codes for the ensemble of inputs it receives, which extract information from this input, and then passes the coded information on to the associative system. The coding stage will attempt to develop codes that are in some sense optimal, both in capturing regularities in the data and in allowing different stimulus inputs to be discriminated. Parallel-distributed processing algorithms that attempt to do this have been available for some time. An example is Rumelhart and Zipser's (1986) competitive-learning algorithm.

Given that this type of *differentiation* approach to perceptual learning is just as straightforward to implement as an associative account, how are we to choose between them? There is a serious problem with the coding system approach: It will necessarily slow down the system's ability to learn associations. If the input to be associated varies in its early stages, while the appropriate codes are being developed, associative learning will be retarded and initially unreliable. Proponents of such a system may argue that the codes developed make subsequent associative learning rapid and more than compensate for any initial problems, but this initial cost may be important if speed of association is vital. A further consideration that follows from this analysis might be termed the stability problem. If elements can be reassigned to different feature sets that best code for the stimulus ensemble currently being experienced, this opens up the possibility that an element that was used to code for one feature after experience with one ensemble might now be used to code for a quite different feature or combination of features after further experience with different stimuli. Any associations formed to that element in its earlier guise will now be, at best, irrelevant or, at worst, wholly inappropriate. It is clearly impossible to devise a successful associative-learning mechanism if the representations to be associated are subject to sudden, unpredictable, and substantial change. This is the stability problem, and models that rely on some coding scheme abstracted on the basis of experience with a stimulus ensemble are always vulnerable to this problem, because fluctuations in stimulus experience will tend to cause fluctuations in the coding schemes adopted by the model.

More recently, Saksida (1999) has implemented a form of neo-Gibsonian perceptual-learning mechanism by further refining the competitive-learning approach given in Rumelhart and Zipser (1986). In her model, units in a second layer compete to code a given pattern of activation in an initial layer of units corresponding to some stimulus input, and the algorithm allows the winning unit (and its near neighbors) in the second, competitive layer to

strengthen their connections in such a fashion as to be more easily activated should that input pattern recur. This has the effect that preexposure to two similar stimuli drives further apart the units in the competitive layer representing those stimuli, so that each is less activated by the other (similar) stimulus input. In effect, more units in the competitive layer become devoted to representing that region of stimulus space, particularly the region between the stimuli, and this increase in representational power ensures that stimuli in that region become more discriminable.

Saksida's (1999) model is in a good position to explain why training on a between-category discrimination can facilitate the learning of a within-category discrimination. As we have noted, Gibson's approach has difficulty with this finding, since, according to her analysis, the features necessary to solve the between-category discrimination should become more salient and impair solution of a subsequent within-category problem. Saksida, on the other hand, is able to argue that exposure to the exemplars of either category will expand the representational space allocated that category, thereby making discriminations within it easier. However, her model seems to fare no better than the basic Gibsonian account when it comes to explaining the critical importance of category structure. As was noted earlier, McLaren (1997) and A. J. Wills and McLaren (1998) have shown that preexposure will only facilitate a within-category discrimination when the exemplars of the category are generated as variations from a prototype. It seems clear that any category containing stimuli similar enough to one another to be grouped together should, on Saksida's account, be susceptible to the perceptual-learning mechanism contained within her model.

Her account also has some difficulty in explaining why preexposure to the category prototype alone should be just as effective in enhancing a within-category discrimination as preexposure to exemplars generated from that prototype (Aitken et al., 1996; Attneave, 1957; the vinegar categorization experiment of Bennett & Mackintosh, in press; see p. 229). In the case in which only a single stimulus from a given category is preexposed, the mechanism for expanding that region of stimulus space does not apply in the same way as before, since there are no discriminations within that region that the network has to make. We can expect that the network would devote many units to representing the preexposed stimulus and that this would lead to those units' also being used to represent similar stimuli. This could, in some circumstances, result in some increase in discriminability owing to the allocation of more representational resources to this region of stimulus space: The distance between winning competitive units representing similar exemplars might increase as a result of more units' being tuned to the preexposed prototype. On the other hand, the distance might decrease if the best available competitive unit (and hence the winner) for each stimulus is now the unit representing the prototype itself (this will depend, among other things, on the amount of noise in the connections). In either case, the model does not predict the same magnitude of perceptual-learning effect observed with exemplar preexposure.[4]

Saksida's (1999) model does, however, unambiguously predict that, if animals are conditioned to AX, generalization to BX will be equally reduced by preexposure either to AX or to BX alone. In her model, preexposure enhances discriminability by increasing the space devoted to the preexposed stimulus, thus ensuring that its representation will overlap less with other, similar stimuli. Preexposure to either AX or BX should thus be equally effective in separating their representations. Our own analysis implies a quite different outcome. We attribute the perceptual learning produced by such preexposure to the latent inhibition of the X elements shared in common by AX and BX. Although preexposure to either AX or BX will result in equal latent inhibition of X, preexposure to AX will also result in latent inhibition of A, whereas preexposure to BX will leave the salience of A elements untouched. Preexposure to BX, therefore, will ensure greater overshadowing of X by A on conditioning trials to AX and will thus be more effective in reducing generalization from AX to BX. Studies of both flavor aversion and appetitive conditioning in rats have confirmed that, even though preexposure to AX results in slower conditioning to AX than does preexposure to BX, it also results in more generalized responding to BX (Honey, 1990; Honey & Hall, 1989).[5]

The question then arises as to whether there is any *need* to postulate the nonassociative differentiation mechanism proposed by Saksida (1999). We acknowledge that there is a certain intuitive plausibility to the idea of perceptual differentiation. And we certainly acknowledge that Saksida's implementation does provide a mechanism for an idea that has usually seemed too nebulous to allow serious discussion. But we remain to be convinced that there is any experimental evidence that requires postulation of such a mechanism.

Some Caveats

The astute reader will doubtless have noted many imperfections in our theorizing. We are not blind to some of these ourselves, and it may be worth acknowledging two classes: first, some cases in which it is difficult to derive unambiguous predictions, and second, some topics and principles that undoubtedly lie outside the scope of our model.

There are numerous situations in which the model postulates tendencies or processes working in opposition to one another and in which the experimental outcome predicted by the model will therefore depend on the balance between these opposing tendencies. We do not see this as a weakness, if only because closely related but slightly different experiments often yield different outcomes. For example, unreinforced preexposure to two stimuli sometimes facilitates and sometimes retards the learning of a discrimination between them. A successful

theory must be able to predict both outcomes and should also be able to point to the critical factor(s) determining which outcome will occur. We suggest that both of these particular outcomes can be explained in terms of latent inhibition. Facilitation will be observed if the effect of the latent inhibition of the elements common to the two stimuli outweighs the effect of latent inhibition of their unique elements. Retardation will be observed when the latter outweighs the former. The most obvious factor determining which outcome will be observed, then, is the relative proportion of common and unique elements, which translates into the difficulty of the discrimination. Even if we cannot make a precise, quantitative prediction here, there is ample evidence to support the qualitative prediction that the more difficult the discrimination, the more likely it is that preexposure will facilitate rather than retard learning (e.g., Chamizo & Mackintosh, 1989; Oswalt, 1972; Trobalon et al., 1991). There is also evidence that directly manipulating the proportion of common elements has the predicted effect (Mackintosh et al., 1991).

A second factor that can influence the outcome of such experiments is whether preexposure and discrimination training occur in the same or different contexts (e.g., Channell & Hall, 1981; Trobalon et al., 1992). In Channell and Hall's experiment, facilitation was observed only following a change of context between preexposure and discrimination training, whereas in Trobalon et al.'s, it was observed only when there was no change of context. Once again, we see it as a virtue of our account that it can explain these conflicting findings and can suggest what may be the critical factor determining one outcome rather than the other—namely, the sheer amount of preexposure (see p. 231). That prediction has not been tested, but although once again it is only qualitative and not quantitative, it certainly could be.

Preexposure to a single stimulus normally retards subsequent conditioning to that stimulus (i.e., the phenomenon of latent inhibition). But there are occasions where conditioning is facilitated by a modest amount of preexposure (e.g., Fanselow, 1990; Kiernan & Westbrook, 1993). We explain these instances of facilitation by our principle of unitization, which then predicts that the critical factor determining the outcome of these experiments is the complexity of the stimulus. We do not, of course, have a precise measure of complexity, although we have provided qualitative support for this prediction (Bennett et al., 1996). The principle of unitization can also predict that, in situations in which preexposure facilitates conditioning to a CS, it may also reduce generalization to another, similar stimulus (Kiernan & Westbrook, 1993). This further prediction depends on the assumption that the sampling mechanism that underlies unitization is not random but may be biased toward features shared in common by the CS and the test stimulus. We acknowledge that we have no formal rules governing the sampling process, but once again, the qualitative prediction seems testable.

No theory of learning can seriously claim to be complete. The central concern of our theorizing has been to explain the phenomena of latent inhibition and perceptual learning within an associative framework. That framework should also apply to a wider range of phenomena, including the basic results of Pavlovian conditioning, discrimination, and generalization (these last will be discussed in a second paper). But we do not pretend to have an explanation for everything. Even within the realm of perceptual learning, it is possible that the process of perceptual differentiation, outlined by E. J. Gibson (1969) and instantiated in Saksida's (1999) network model, plays a part. We do not know of any evidence that *requires* the postulation of such a process, but absence of evidence is not the same as evidence of absence.

In other areas of conditioning and discrimination learning, we freely acknowledge that there is evidence of absence. At least some (but by no means all) of the lacunae in the Rescorla–Wagner (1972) model, identified by Miller, Barnet, and Grahame (1995), apply to our model also. We have no elaborate performance rule, and if some version of Miller and colleagues' comparator hypothesis (Miller & Matzel, 1988) were to become generally accepted, we should need to acknowledge that. Our account of latent inhibition treats it as a loss of salience or associability, not as a failure of performance or retrieval or as associative interference. To the extent that there is evidence consistent with these alternative formulations, this too lies outside the scope of the model. We should be happy to acknowledge that many of the phenomena of interest to the learning theorist are multiply determined.

Our account of latent inhibition bears a certain resemblance to that provided by SOP (Wagner, 1981). But we should also acknowledge that there are quite certainly other principles governing changes in associability. Mackintosh (1975) and Pearce and Hall (1980) have both proposed a further mechanism, the former on the basis of the relative predictive value of a target stimulus, the latter on the predictability of the outcome following the target. Both of these formulations can predict the results of certain experiments on blocking that we are unable to explain (see, e.g., Dickinson, Hall, & Mackintosh, 1976; Mackintosh, Bygrave, & Picton, 1977). Mackintosh's account receives specific support from demonstrations of the reinforcer specificity of changes in associability (e.g., Bennett, Wills, Oakeshott, & Mackintosh, 2000; Dickinson & Mackintosh, 1979), whereas Pearce and Hall's central postulate is supported by direct evidence that preexposure to a stimulus followed by a relatively unpredictable outcome maintains its associability (e.g., Pearce, Kaye, & Hall, 1982).

Theories of selective attention in discrimination learning (e.g., Mackintosh, 1975; Sutherland & Mackintosh, 1971) can also point to evidence of positive transfer effects between discriminations sharing similar relevant cues that lie outside the scope of the model proposed here. To take one example that we do address, search

image effects (p. 233) have traditionally been explained by the proposal that predators learn to attend to the distinctive features of an abundant prey and, since attention is highly selective, this interferes with their ability to attend to the distinctive features of less abundant prey. However, Plaisted (1997) demonstrated that the entire search image effect observed in her experiments was sufficiently explained by noting that the average interval of time separating the occurrence of two instances of a less frequent target was greater than that separating two instances of a more frequent target. Her data required only the postulation of trace decay, rather than any appeal to selective attention. But that might be because her two targets were relatively similar and would possibly have both been detected by attention to the same class of features. If one target differed from its background in color and the other in pattern, it is possible that this would have engaged an attentional mechanism.

CONCLUSIONS

We conclude that elemental analyses of associative learning still have a lot to offer. The model we proposed 10 years ago has provided a more powerful, and readily testable, account of the phenomena of perceptual learning than was hitherto available. No doubt, data will emerge, even in this restricted domain, that will require modifications to our analysis or even suggest alternative explanations. But we are not aware of any evidence today that is inconsistent with our analysis. In other areas, as we have acknowledged, our account is almost certainly incomplete. But at the very least, we are a long way from understanding the limits of an elemental approach or the conditions that will require a configural account. Our analysis suggests that an elemental component is required in any theory of associative learning and that such a component will be both a substantial and a significant part of the theory.

REFERENCES

Aitken, M. R. F., Bennett, C. H., McLaren, I. P. L., & Mackintosh, N. J. (1996). Perceptual differentiation during categorization learning by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **22**, 43-50.

Atkinson, R. C., & Estes, W. K. (1963). Stimulus sampling theory. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 2, pp. 121-268). New York: Wiley.

Attneave, F. (1957). Transfer of experience with a class-schema to identification-learning of patterns and shapes. *Journal of Experimental Psychology*, **54**, 81-88.

Baker, A. G. (1974). Conditioned inhibition is not the symmetrical opposite of conditioned excitation: A test of the Rescorla–Wagner model. *Learning & Motivation*, **5**, 369-379.

Baker, A. G., & Mercier, P. (1982). Extinction of the context and latent inhibition. *Learning & Motivation*, **13**, 391-416.

Bennett, C. H., & Mackintosh, N. J. (in press). Comparison and contrast as a mechanism of perceptual learning. *Quarterly Journal of Experimental Psychology*.

Bennett, C. H., Scahill, V. L., Griffiths, D. P., & Mackintosh, N. J. (1999). The role of inhibitory associations in perceptual learning. *Animal Learning & Behavior*, **27**, 333-345.

Bennett, C. H., Tremain, M., & Mackintosh, N. J. (1996). Facilitation and retardation of flavour aversion conditioning following prior exposure to the CS. *Quarterly Journal of Experimental Psychology*, **49B**, 220-230.

Bennett, C. H., Wills, S. J., Oakeshott, S. M., & Mackintosh, N. J. (2000). Is the context specificity of latent inhibition a sufficient explanation of learned irrelevance? *Quarterly Journal of Experimental Psychology*, **53B**, 239-254.

Bennett, C. H., Wills, S. J., Wells, J. O., & Mackintosh, N. J. (1994). Reduced generalization following preexposure: Latent inhibition of common elements or a difference in familiarity? *Journal of Experimental Psychology: Animal Behavior Processes*, **20**, 232-239.

Best, M. R., & Batson, J. D. (1977). Enhancing the expression of flavor neophobia: Some effects of the ingestion–illness contingency. *Journal of Experimental Psychology: Animal Behavior Processes*, **3**, 132-143.

Bevins, R. A., & Ayres, J. J. B. (1995). One-trial context fear conditioning as a function of the interstimulus interval. *Animal Learning & Behavior*, **23**, 400-410.

Bevins, R. A., McPhee, J. E., Rauhut, A. S., & Ayres, J. J. B. (1997). Converging evidence for one-trial context fear conditioning with an immediate shock: Importance of shock potency. *Journal of Experimental Psychology: Animal Behavior Processes*, **23**, 312-324.

Blaisdell, A. P., Denniston, J. C., & Miller, R. R. (1998). Temporal encoding as a determinant of overshadowing. *Journal of Experimental Psychology: Animal Behavior Processes*, **24**, 72-83.

Blanchard, R. J., Fukunaga, K. K., & Blanchard, D. C. (1976). Environmental control of defensive reactions to footshock. *Bulletin of the Psychonomic Society*, **8**, 129-130.

Bouton, M. L. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, **114**, 80-99.

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review*, **58**, 313-323.

Carr, A. F. (1974). Latent inhibition and overshadowing in conditioned emotional response conditioning in rats. *Journal of Comparative & Physiological Psychology*, **86**, 718-723.

Chamizo, V. D., & Mackintosh, N. J. (1989). Latent learning and latent inhibition in maze discriminations. *Quarterly Journal of Experimental Psychology*, **41B**, 21-31.

Channell, S., & Hall, G. (1981). Facilitation and retardation of discrimination learning after exposure to the stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, **7**, 437-446.

Cole, R. P., Barnet, R. C., & Miller, R. R. (1995). Temporal encoding in trace conditioning. *Animal Learning & Behavior*, **23**, 144-153.

Darby, R. J., & Pearce, J. M. (1997). The effect of stimulus preexposure on responding during a compound stimulus. *Quarterly Journal of Experimental Psychology*, **50B**, 203-216.

Davis, M. (1970). Effects of interstimulus interval length and variability on startle-response habituation in the rat. *Journal of Comparative & Physiological Psychology*, **72**, 177-192.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgments. *Quarterly Journal of Experimental Psychology*, **48B**, 60-80.

Dickinson, A., Hall, G., & Mackintosh, N. J. (1976). Surprise and the attenuation of blocking. *Journal of Experimental Psychology: Animal Behavior Processes*, **2**, 313-322.

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

Dwyer, D. M. (1999). Retrospective revaluation or mediated conditioning? The effect of different reinforcers. *Quarterly Journal of Experimental Psychology*, **52B**, 289-306.

Dwyer, D. M., Mackintosh, N. J., & Boakes, R. A. (1998). Simultaneous activation of the representation of absent cues result in the strengthening of an excitatory association between them. *Journal of Experimental Psychology: Animal Behavior Processes*, **24**, 163-171.

Elkins, R. L. (1973). Attenuation of drug-induced bait-shyness to a palatable solution as an increasing function of its availability prior to conditioning. *Behavioral Biology*, **9**, 221-226.

Ellis, W. R., III (1970). Role of stimulus sequences in stimulus dis-

crimination and stimulus generalization. *Journal of Experimental Psychology*, **83**, 155-163.

Espinet, A., Iraola, J. A., Bennett, C. H., & Mackintosh, N. J. (1995). Inhibitory associations between neutral stimuli in flavor-aversion conditioning. *Animal Learning & Behavior*, **23**, 361-368.

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, **62**, 369-377.

Estes, W. K. (1959). The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 2, pp. 380-491). New York: McGraw-Hill.

Fanselow, M. S. (1986). Associative vs. topographical accounts of the immediate shock freezing deficit in rats: Implications for the response selection rules governing species specific defensive reactions. *Learning & Motivation*, **17**, 16-39.

Fanselow, M. S. (1990). Factors governing one-trial contextual conditioning. *Animal Learning & Behavior*, **18**, 264-270.

Fenwick, S., Mikulka, P. J., & Klein, S. B. (1975). The effect of different levels of pre-exposure to sucrose on the acquisition and extinction of a conditioned aversion. *Behavioral Biology*, **14**, 231-235.

Forgus, R. H. (1958a). The effect of different kinds of form preexposure on form discrimination learning. *Journal of Comparative & Physiological Psychology*, **51**, 75-78.

Forgus, R. H. (1958b). The interaction between form preexposure and test requirements in determining form discrimination. *Journal of Comparative & Physiological Psychology*, **51**, 588-591.

Fudim, O. K. (1978). Sensory preconditioning of flavors with a formalin-induced sodium need. *Journal of Experimental Psychology: Animal Behavior Processes*, **4**, 276-285.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts.

Gibson, E. J., & Levin, H. (1975). *The psychology of reading*. Cambridge, MA: MIT Press.

Gibson, E. J., & Walk, R. D. (1956). The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *Journal of Comparative & Physiological Psychology*, **49**, 239-242.

Gibson, E. J., Walk, R. D., Pick, H. L., & Tighe, T. J. (1958). The effect of prolonged exposure to visual patterns on learning to discriminate similar and different patterns. *Journal of Comparative & Physiological Psychology*, **51**, 584-587.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review*, **62**, 32-41.

Grahame, N. J., Barnet, R. C., Gunther, L. M., & Miller, R. R. (1994). Latent inhibition as a performance deficit resulting from CS–context associations. *Animal Learning & Behavior*, **22**, 395-408.

Gulliksen, H., & Wolfle, H. L. (1938). A theory of learning and transfer: I. *Psychometrika*, **3**, 127-149.

Hall, G. (1980). Exposure learning in animals. *Psychological Bulletin*, **88**, 535-550.

Hall. G. (1991). *Perceptual and associative learning.* Oxford: Oxford University Press, Clarendon Press.

Hall, G., & Channell, S. (1985). Differential effects of contextual change on latent inhibition and on the habituation of an orienting response. *Journal of Experimental Psychology: Animal Behavior Processes*, **11**, 470-481.

Hall, G., & Channell, S. (1986). Context specificity of latent inhibition in taste aversion learning. *Quarterly Journal of Experimental Psychology*, **38B**, 121-139.

Hall, G., & Minor, H. (1984). A search for context–stimulus associations in latent inhibition. *Quarterly Journal of Experimental Psychology*, **36B**, 145-169.

Hebb, D. O. (1949). *Organization of behavior.* New York: Wiley.

Holland, P. C. (1981). Acquisition of representation-mediated conditioned food aversions. *Learning & Motivation*, **12**, 1-18.

Honey, R. C. (1990). Stimulus generalization as a function of stimulus novelty and familiarity in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, **16**, 178-184.

Honey, R. C., & Bateson, P. (1996). Stimulus comparison and perceptual learning: Further evidence and evaluation from an imprinting procedure. *Quarterly Journal of Experimental Psychology*, **49B**, 259-269.

Honey, R. C., Bateson, P., & Horn, G. (1994). The role of stimulus comparison in perceptual learning: An investigation with the domestic chick. *Quarterly Journal of Experimental Psychology*, **47B**, 83-103.

Honey, R. C., & Hall, G. (1989). Enhanced discriminability and reduced associability following flavor preexposure. *Learning & Motivation*, **20**, 262-277.

Hull, C. L. (1943). *Principles of behavior.* New York: Appleton-Century-Crofts.

James, J. H., & Wagner, A. R. (1980). One-trial overshadowing: Evidence of distributive processing. *Journal of Experimental Psychology: Animal Behavior Processes*, **6**, 188-205.

James, W. (1890). *Principles of psychology.* New York: Holt.

Jones, F., Wills, A. J., & McLaren, I. P. L. (1998). Perceptual categorisation: Connectionist modeling and decision rules. *Quarterly Journal of Experimental Psychology*, **51B**, 33-58.

Kiernan, M. J., & Westbrook, R. F. (1993). Effects of exposure to a to-be-shocked environment upon the rat's freezing response: Evidence for facilitation, latent inhibition, and perceptual learning. *Quarterly Journal of Experimental Psychology*, **46B**, 271-288.

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

Leonard, S., & Hall, G. (in press). Inhibitory associations between neutral stimuli? A test using the conditioned suppression procedure. *Quarterly Journal of Experimental Psychology*.

Lovejoy, E. (1968). *Attention in discrimination learning.* San Francisco: Holden-Day.

Lovibond, P., Preston, G. C., & Mackintosh, N. J. (1984). Contextual control of conditioning and latent inhibition. *Journal of Experimental Psychology: Animal Behavior Processes*, **10**, 360-375.

Mackintosh, N. J. (1974). *The psychology of animal learning*. New York: Academic Press.

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

Mackintosh, N. J. (1983). *Conditioning and associative learning.* Oxford: Oxford University Press.

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

Mackintosh, N. J., Kaye, H., & Bennett, C. H. (1991). Perceptual learning in flavour aversion conditioning. *Quarterly Journal of Experimental Psychology*, **43B**, 297-322.

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

Marlin, N. A., & Miller, R. R. (1981). Associations to contextual stimuli as a determinant of long-term habituation. *Journal of Experimental Psychology: Animal Behavior Processes*, **7**, 313-333.

Matzel, L. D., Schactman, T. R., & Miller, R. R. (1985). Learned irrelevance exceeds the sum of CS-preexposure and US-preexposure deficits. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 311-319.

McCaslin, E. F. (1954). Successive and simultaneous discrimination as a function of stimulus similarity. *American Journal of Psychology*, **67**, 308-314.

McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, **114**, 159-188.

McLaren, I. P. L. (1997). Categorisation and perceptual learning: An analogue of the face inversion effect. *Quarterly Journal of Experimental Psychology*, **50A**, 257-273.

McLaren, I. P. L., Bennett, C. H., Plaisted, K. C., Aitken, M. R. F., & Mackintosh, N. J. (1994). Latent inhibition, context specificity, and context familiarity. *Quarterly Journal of Experimental Psychology*, **47B**, 387-400.

McLaren, I. P. L., Kaye, H., & Mackintosh, N. J. (1989). An associative theory of the representation of stimuli: Applications to perceptual learning and latent inhibition. In R. G. M. Morris (Ed.), *Parallel distributed processing: Implications for psychology and neurobiology* (pp. 102-130). Oxford: Oxford University Press, Clarendon Press.

McLaren, I. P. L., Leevers, H. L., & Mackintosh, N. J. (1994). Recognition, categorisation and perceptual learning. In C. Umiltà & M. Moscovitch (Eds.), *Attention and performance XV: Conscious and nonconscious information processing* (pp. 889-909). Cambridge, MA: MIT Press.

Medin, D. L. (1975). A theory of context in discrimination learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 263-314). New York: Academic Press.

Miller, R. R., Barnet, R. C., & Grahame, N. J. (1995). Assessment of the Rescorla–Wagner model. *Psychological Bulletin*, **117**, 363-386.

Miller, R. R., Kasprow, W. J., & Schachtman, T. R. (1986). Retrieval variability: Sources and consequences. *American Journal of Psychology*, **99**, 145-218.

Miller, R. R., & Matute, H. (1996). Biological significance in forward and backward blocking: Resolution of a discrepancy between animal conditioning and human causal judgment. *Journal of Experimental Psychology: General*, **125**, 370-386.

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 51-92). San Diego: Academic Press.

Navarro, J. I., Hallam, S. C., Matzel, L. D., & Miller, R. R. (1989). Superconditioning and overshadowing. *Learning & Motivation*, **20**, 130-152.

Oswalt, R. M. (1972). Relationship between level of visual pattern difficulty during rearing and subsequent discrimination in rats. *Journal of Comparative & Physiological Psychology*, **81**, 122-125.

Pearce, J. M. (1987). A model of stimulus generalization for Pavlovian conditioning. *Psychological Review*, **94**, 61-73.

Pearce, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, **101**, 587-607.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, **87**, 532-552.

Pearce, J. M., Kaye, H., & Hall, G. (1982). Predictive accuracy and stimulus associability: Development of a model for Pavlovian learning. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analysis of behavior: Acquisition* (pp. 241-256). Cambridge, MA: Ballinger.

Pietrewicz, A. T., & Kamil, A. C. (1979). Search image formation in the blue jay *(Cyanicitta cristata)*. *Science*, **204**, 1332-1333.

Plaisted, K. C. (1997). The effect of interstimulus interval on the discrimination of cryptic targets. *Journal of Experimental Psychology: Animal Behavior Processes*, **23**, 248-259.

Plaisted, K. C., & Mackintosh, N. J. (1995). Visual search for cryptic stimuli in pigeons: Implications for the search image and search rate hypothesis. *Animal Behavior*, **50**, l2l9-l232.

Poggio, T., Fahle, M., & Edelman, S. (1992). Fast perceptual learning in visual hyperacuity. *Science*, **256**, 1018-1021.

Prados, J., Chamizo, V. D., & Mackintosh, N. J. (1999). Latent inhibition and perceptual learning in a swimming pool navigation task. *Journal of Experimental Psychology: Animal Behavior Processes*, **25**, 37-44.

Rescorla, R. A., & Durlach, P. J. (1981). Within-event learning in Pavlovian conditioning. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 81-111). Hillsdale, NJ: Erlbaum.

Rescorla, R. A., & Wagner, A. R. (1972). 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* (pp. 64-99). New York: Appleton-Century-Crofts.

Riley, D. A. (1958). The nature of the effective stimulus in animal discrimination learning: Transposition reconsidered. *Psychological Review*, **65**, 1-7.

Rodrigo, T., Chamizo, V. D., McLaren, I. P. L., & Mackintosh, N. J. (1994). Effects of preexposure to the same or different pattern of extra-maze cues on subsequent extra-maze discrimination. *Quarterly Journal of Experimental Psychology*, **47B**, 15-26.

Rodrigo, T., Chamizo, V. D., McLaren, I. P. L., & Mackintosh, N. J. (1997). Blocking in the spatial domain. *Journal of Experimental Psychology: Animal Behavior Processes*, **23**, 110-118.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations* (pp. 318-362). Cambridge, MA: MIT Press, Bradford Books.

Rumelhart, D. E., & Zipser, D. (1986). Feature discovery by competitive learning. In D. E. Rumelhart, J. L. McClelland, & the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 1: Foundations*. Cambridge, MA: MIT Press, Bradford Books.

Saksida, L. M. (1999). Effects of similarity and experience on discrimination learning: A nonassociative connectionist model of perceptual learning. *Journal of Experimental Psychology: Animal Behavior Processes*, **25**, 308-323.

Saldanha, E. L., & Bitterman, M. E. (1951). Relational learning in the rat. *American Journal of Psychology*, **64**, 37-53.

Sansa, J., Chamizo, V. D., & Mackintosh, N. J. (1996). Apprendizaje perceptivo en discriminaciones espaciales. *Psicología*, **17**, 279-295.

Schnur, P., & Lubow, R. E. (1976). Latent inhibition: The effects of ITI and CS intensity during preexposure. *Learning & Motivation*, **7**, 540-550.

Spence, K. W. (1952). The nature of the response in discrimination learning. *Psychological Review*, **59**, 89-93.

Sutherland, N. S., & Mackintosh, N. J. (1971). *Mechanisms of animal discrimination learning.* New York: Academic Press.

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, **88**, 135-170.

Swan, J. A., & Pearce, J. M. (1988). The orienting response as an index of stimulus associability in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, **14**, 292-301.

Symonds, M., & Hall, G. (1995). Perceptual learning in flavor aversion conditioning: Roles of stimulus comparison and latent inhibition of common elements. *Learning & Motivation*, **26**, 203-219.

Symonds, M., & Hall, G. (1997). Stimulus preexposure, comparison, and changes in the associability of common stimulus features. *Quarterly Journal of Experimental Psychology*, **50B**, 317-331.

Thompson, R. F. (1965). The neural basis of stimulus generalization. In D. I. Mostofsky (Ed.), *Stimulus generalization* (pp. 154-178). Stanford: Stanford University Press.

Thorndike, E. L. (1911). *Animal intelligence: Experimental studies.* New York: Macmillan.

Tinbergen, L. (1960). The natural control of insects in pinewoods. Factors influencing the intensity of predation by songbirds. *Archives Neerlandaise de Zoologie*, **13**, 265-343.

Trobalon, J. B., Chamizo, V. D., & Mackintosh, N. J. (1992). Role of context in perceptual learning in maze discriminations. *Quarterly Journal of Experimental Psychology*, **44B**, 57-73.

Trobalon, J. B., Sansa, J., Chamizo, V. D., & Mackintosh, N. J. (1991). Perceptual learning in maze discriminations. *Quarterly Journal of Experimental Psychology*, **43B**, 389-402.

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning & Motivation*, **25**, 127-151.

Wagner, A. R. (1978). Expectancies and the priming of STM. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (pp. 177-209). Hillsdale, NJ: Erlbaum.

Wagner, A. R. (1979). Habituation and memory. In A. Dickinson & R. A. Boakes (Eds.), *Mechanisms of learning and motivation* (pp. 53-82). Hillsdale, NJ: Erlbaum.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Infor-*

mation processing in animals: Memory mechanisms (pp. 95-128). Hillsdale, NJ: ErIbaum.

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

Wagner, A. R., & Rescorla, R. A. (1972). Inhibition in Pavlovian conditioning: Application of a theory. In R. A. Boakes & M. S. Halliday (Eds.), *Inhibition and learning* (pp. 301-336). London: Academic Press.

Ward-Robinson, J., & Hall, G. (1996). Backward sensory preconditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, **22**, 395-404.

Westbrook, R. F., Bond, N. W., & Feyer, A. M. (1981). Short-term and long-term decrements in toxicosis-induced odor-aversion learning: The role of duration of exposure to an odor. *Journal of Experimental Psychology: Animal Behavior Processes*, **7**, 362-381.

Widrow, G., & Hoff, M. E. (1960). Adaptive switching circuits. *Institute of Radio Engineers: Western electronic show and convention* (Convention Record, Pt. 4), pp. 96-104.

Wills, A. J., & McLaren, I. P. L. (1997a). Generalisation in human category learning: a connectionist explanation of discriminative versus non-discriminative training gradient differences. *Quarterly Journal of Experimental Psychology*, **50A**, 607-630.

Wills, A. J., & McLaren, I. P. L. (1997b). Generalisation in human category learning: One and two category problems. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Wills, A. J., & McLaren, I. P. L. (1998). Perceptual learning and free classification. *Quarterly Journal of Experimental Psychology*, **51B**, 235-270.

Wills, S. J., & Mackintosh, N. J. (1999). Relational learning in pigeons? *Quarterly Journal of Experimental Psychology*, **52B**, 31-52.

Zeaman, D., & House, B. J. (1963). The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency: Psychological theory and research* (pp. 159-223). New York: McGraw-Hill.

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

NOTES

1. Our principled answer to this would be to appeal to the type of connectionist decision mechanism discussed in Jones, Wills, and McLaren (1998) and A. J. Wills and McLaren (1997a, 1997b). In particular, see A. J. Wills and McLaren, 1997b, for an example of how decision processes can produce effects that might be thought attributable to more basic learning mechanisms.

2. Taken by themselves, these differences in consumption of BX might be attributable to differences in neophobia, since Groups Same and Diff were both preexposed to BX, whereas the control group was not. However, there is evidence that habituation, unlike latent inhibition, is not context specific (e.g., Hall & Channell, 1985; Marlin & Miller, 1981), so this suggestion cannot explain the difference between Groups Same and Diff. Moreover, other studies from our laboratory employing identical procedures and flavors have found no evidence of neophobia's being attenuated by a single preexposure (Bennett et al., 1994).

3. In these experiments, the precise location of the cryptic target varies at random from trial to trial, and/or it is presented against a background that itself varies from trial to trial. This will work against the formation of associations between target and background, which is a process that would otherwise lead to poor discrimination by eliciting false-positive responses.

4. Our own explanation of all these findings appeals to latent inhibition of common elements. Although part of Saksida's (1999) general model incorporates a mechanism for latent inhibition, she stresses that a unique aspect of her model is that "it suggests that perceptual learning is based on a singular, nonassociative mechanism that focuses on the separation of the representations of stimuli" (p. 319). A brief, cryptic qualification of this claim does not, in our view, acknowledge the central role played by latent inhibition in generating many instances of perceptual learning.

5. There was no explicitly manipulated common feature, X, in these experiments, but the A and B stimuli used were clearly sufficiently similar to allow substantial generalization between them.

(Manuscript received January 12, 2000;
revision accepted for publication April 19, 2000.)