

# Evaluation and development of a connectionist theory of configural learning

JOHN M. PEARCE  
*Cardiff University, Cardiff, Wales*

A configural theory of associative learning is described that is based on the assumption that conditioning results in associations between the unconditioned stimulus and a representation of the entire pattern of stimulation that was present prior to its delivery. Configural theory was formulated originally to account for generalization and discrimination in Pavlovian conditioning. The first part of the article demonstrates how this theory can be used to explain results from studies of overshadowing, blocking, summation, and discrimination learning. The second part of the article shows how the theory can be developed to explain a broader range of phenomena, including mediated conditioning, reinforcer devaluation effects, the differential outcomes effect, acquired equivalence, sensory preconditioning, and structural discriminations.

Theories of associative learning are built on the assumption that the repeated pairing of two events in succession will result in the growth of a connection, or association, between their internal representations. The existence of this association will then allow the first event to excite the representation of the second event, which may then lead to the performance of a response. The change in the way an animal reacts to a conditioned stimulus (CS) as a result of its being paired with an unconditioned stimulus (US) provides one example of the way in which association formation is believed to influence behavior. An important difference among theories of associative learning concerns the assumptions they make about the associations that are formed when the first event is a compound of two or more elements. According to what will be referred to as *elemental* theories, this type of training will provide the opportunity for each element of the compound to enter into an association with the representation of the second event. In contrast, what will be referred to as *configural* theories are based on the principle that a representation of the entire pattern of stimulation that constitutes the first event will be formed and will enter into a single association with the second event. This difference has significant implica-

tions for the way in which the two classes of associative theory are developed formally and for the predictions they make.

Historically, elemental theories have proved more popular than configural theories. The theories of Hull (1943, 1952) and Spence (1936, 1937, 1952) were based on the assumption that conditioning encouraged the growth of stimulus-response connections between individual stimuli and the response that occurred in their presence. More recently, the most influential theory of associative learning has been the one proposed by Rescorla and Wagner (1972; Wagner & Rescorla, 1972). The elemental nature of this theory is made evident by the assumption that a compound conditioning trial provides the opportunity for a change in the associative strength of each element of the compound. A number of associative theories that have been proposed subsequent to the Rescorla-Wagner theory differ from it in various ways, but the majority share its elemental basis (Kehoe, 1988; Mackintosh, 1975; McLaren, Kaye, & Mackintosh, 1989; Pearce & Hall, 1980; Wagner, 1976, 1978, 1981).

In contrast, theories that have adopted a configural approach to conditioning have been rather few and have had less of an impact than have their elemental counterparts. Gulliksen and Wolfe (1938a, 1938b) developed a formal theory that was used to explain the way in which animals solve simultaneous discriminations between stimuli that differ on a single dimension. Suppose that an animal must approach a black door and avoid a white door, irrespective of their positions, in order to gain food; it follows from the proposals of Gulliksen and Wolfe (1938a, 1938b) that they will acquire a specific response to each configuration to which they are exposed—for example, go left, to the pattern of black door to the left of the white door. The mixed model of conditioning described by Atkinson and Estes (1963, p. 243) can also be regarded as a configural theory (see also Estes & Hopkins, 1961; Friedman & Gelfand,

---

The research described in this article was supported by grants from the U.K. Biotechnology and Biological Sciences Research Council and the Medical Research Council. The article was written while I was at the Institute for Advanced Study at the University of Indiana, Bloomington, and an Erskine Fellow at the University of Canterbury, Christchurch, New Zealand. I thank both universities for their support and for their generous provision of facilities. I am grateful to Aydan Aydin, David George, Edward Redhead, and Paul Wilson for their many helpful suggestions concerning the work discussed in this article. Correspondence concerning this article should be addressed to J. M. Pearce, School of Psychology, Cardiff University, Cardiff CF1 3YG, Wales (e-mail: pearcejm@cardiff.ac.uk).

—This article was invited by the editors.

1964). According to this model, subjects will select a sample of elements from the stimuli that are presented on a trial, and the entire sample will then enter, as a unit, into an association with the outcome of the trial. More recently, Bellingham, Gillette-Bellingham, and Kehoe (1985) have demonstrated how a configural version of the Rescorla-Wagner (1972) model can be used to explain the way in which animals solve patterning discriminations. Finally, Pearce (1987) has proposed a formal configural theory of associative learning, which has been applied to a considerably wider range of findings than have its predecessors. This theory was based on an informal account of configural learning proposed by Young and Pearce (1984) and was presented as a connectionist network by Pearce (1994).

One purpose of the present article is to assess the relative merits of configural and elemental theories of associative learning. A second purpose is to consider several ways in which configural theory might be elaborated. To date, configural theory has been applied principally to conventional studies of conditioning and discrimination learning, with the result that there are a number of findings that lie beyond the scope of the theory. These include sensory preconditioning, acquired equivalence, and the ability of animals to solve what have been called *structural discriminations*. Thus, ways in which the theory might be developed to explain these and related findings will be considered.

The distinction between configural and elemental theories of learning is not confined to the analysis of research with animals. This distinction can also be found in theories that have been developed to account for how humans solve categorization problems. Gluck and Bower (1988) have developed a connectionist network, based on the Rescorla-Wagner (1972) equation, that has been used to explain some of the outcomes of categorization studies in humans. More complex elemental theories of categorization are provided by, for example, certain connectionist networks that contain a layer of hidden units (e.g., Rumelhart, Hinton, & Williams, 1988). In contrast, an account of categorization based on configural principles can be found in the theories of Hintzman (1986) and Kruschke (1992). Thus, the following discussion is of relevance to the analysis of learning both in animals and in humans.

## OVERVIEW OF CONFIGURAL THEORY

A useful way to introduce the configural theory of Pearce (1987, 1994) is to contrast it with the Rescorla-Wagner (1972) theory. Figure 1A shows a simple connectionist network with two input units, which are activated by Stimuli A and B, and an output unit that is activated by a US. When a stimulus is presented to the network, it will activate its input unit, which will activate the output unit to a degree that is determined by the strength of the connection between them. The Rescorla and Wagner rule stipulates that the change in the strength of the connection be-

tween the input unit, A, and the US unit,  $\Delta V_A$ , on any trial for which A is present is given by Equation 1:

$$\Delta V_A = \alpha\beta(\lambda - V_T). \quad (1)$$

In this equation,  $\lambda$  refers to the asymptote of conditioning, and  $V_T$  is the combined associative strength of all the stimuli that are present on that trial. The value of  $\alpha$  is dictated by the properties of the CS, and the value of  $\beta$  is dictated by the properties of the US. Both of these parameters have values between 0 and 1. The strength of the conditioned response (CR) to two or more stimuli is determined by the sum of the strengths of their connections with the US unit. The elemental nature of the theory should be evident from Figure 1, because each input unit can enter into an association with the US unit. As it has been presented here, the Rescorla-Wagner (1972) model is unable to explain how animals solve certain discriminations, such as negative patterning, in which the US is presented after A and B when they occur by themselves, but not when they occur in compound, A+ B+ AB- (Woodbury, 1943). On the basis of the principles that have just been put forward, the theory predicts that responding will always be stronger in the presence of the compound than in the presence of the individual stimuli and that the discrimination will never be solved. To explain the ability

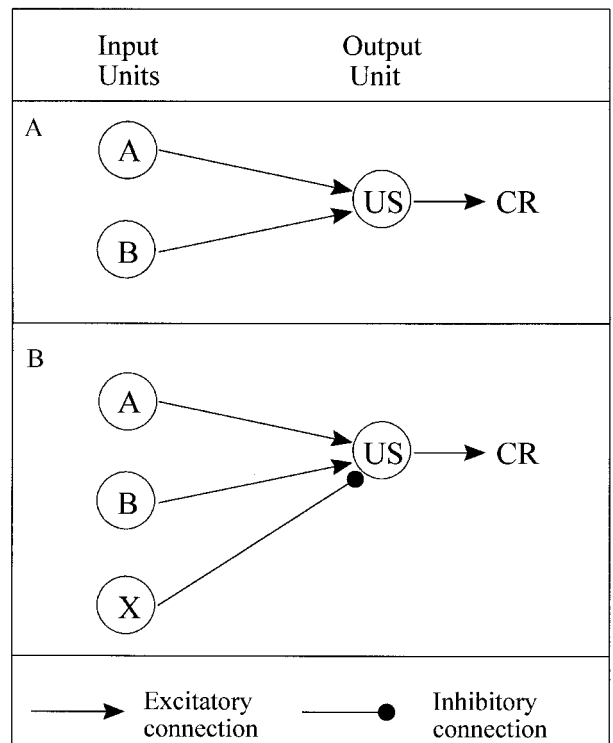


Figure 1. The connections that will develop, according to an elemental theory of conditioning, (A) during conditioning with a compound, AB, and (B) during a negative-patterning discrimination.

of animals to solve this type of discrimination, Wagner and Rescorla (1972) suggested that whenever a compound of two or more stimuli is presented, it will activate additional stimulus element representations that are unique to the compound. These unique cues are able to take part in conditioning in the same way as conventional stimuli. Thus, if Compound AB is paired with a US, the connections that can develop are those shown in Figure 1B, where X represents the unique cue created by the compound, which will be effective only when A and B are presented together. Given a negative-patterning discrimination, Equation 1 predicts that X will acquire negative associative strength and will have an inhibitory influence on responding. Eventually, the magnitude of this influence will be sufficient to counter entirely the excitatory strength acquired by A and B and will result in little or no respond-

ing on the trials with AB. However, when A or B is presented alone, X will not be activated, and the strength of the CR will be determined by the associative strength of the CS.

Figure 2A shows the connections that are predicted to develop by configural theory when a compound, AB, is paired with a US. In essence, the network is a development of the one shown in Figure 1, with an additional, configural unit interposed between the input and the output units. When AB is first presented to the network, the two input units will rapidly become connected to a configural unit. Thereafter, whenever the compound AB is presented, it will activate fully the AB configural unit. If the compound AB should be followed by a US, a connection will develop between the AB configural unit and the US unit. The strength of this connection,  $V_{AB}$ , will develop gradually over trials, in accord with Equation 2:

$$\Delta V_{AB} = \beta(\lambda - V_{AB}). \quad (2)$$

According to Pearce (1994), the strength of the CR predicted on any trial is determined by the level of activation of the configural unit for AB, multiplied by  $V_{AB}$ . If AB is presented, the unit will be activated to its maximal value of 1, but if only one stimulus is presented, the level of activation of the output unit will be less than that if AB had been presented. Equation 3 shows how the strength of the CR in the presence of A,  $E_A$ , is determined, where  ${}_A S_{AB}$  is the similarity of A to AB:

$$E_A = {}_A S_{AB} \cdot V_{AB}. \quad (3)$$

The value of  ${}_A S_{AB}$  can be derived from Equation 4, where P1 and P2 would refer to the two patterns, A and AB.  $N_C$  is the number of input units shared by the two patterns (one, for the present example), and  $N_{P1}$  and  $N_{P2}$  are the number of input elements activated by the patterns (one and two, for the present example):

$${}_A S_{AB} = N_C / N_{P1} \cdot N_C / N_{P2}. \quad (4)$$

It is worth considering a little further the effects of presenting A after conditioning with AB. When A is presented, in addition to activating the AB configural unit, it will become connected to its own configural unit, which will be fully activated whenever A occurs by itself. If A should then be paired with a US, even though two configural units will be activated, the only configural unit to enter into an association with the US will be the fully activated unit for A. The connections that will develop in the network if A is paired with the US are shown in Figure 2B. To take account of the fact that a CR will occur on the first presentation of A, Equation 5 shows the way in which conditioning with A will progress:

$$\Delta V_A = \beta[\lambda - ({}_A S_{AB} \cdot V_{AB} + V_A)]. \quad (5)$$

In this equation, it is not only the associative strength of the configural unit for A that determines the increment in the strength of its association with the output unit, but also

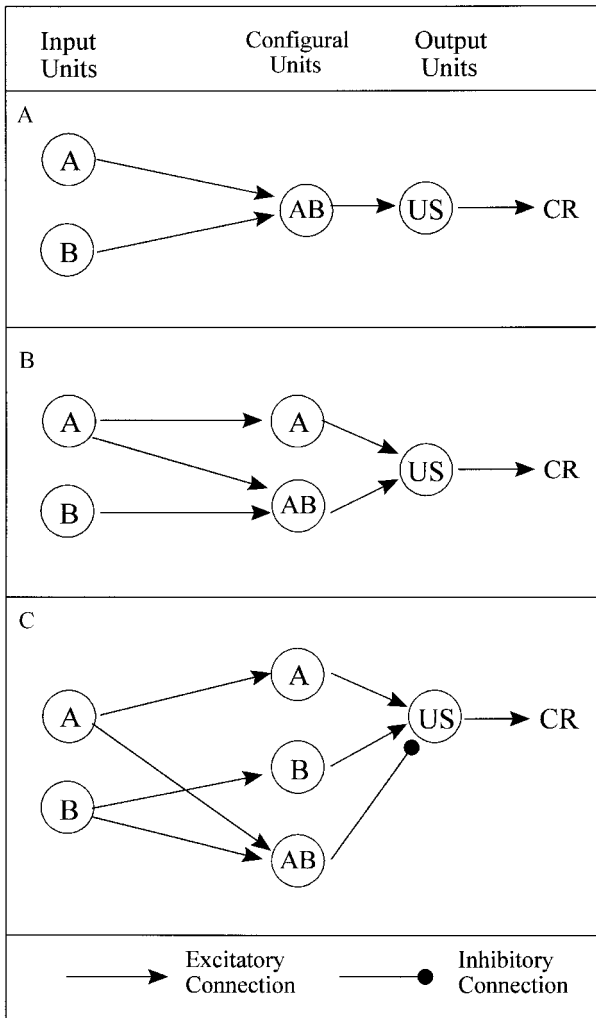


Figure 2. The connections that will develop, according to the configural theory of Pearce (1994), (A) during conditioning with a compound, AB, (B) during conditioning with AB followed by conditioning with A, and (C) during a negative-patterning discrimination.

the extent to which the presence of A is able to excite a CR through activation of the configural unit for AB.

Figure 2C shows the connections that are predicted to be formed when animals are trained with a negative-patterning discrimination. The nonreinforced trials with AB will result in the development of an inhibitory connection between the AB configural unit and the US unit. Thus, when A is presented, it will activate strongly its own configural unit and will activate weakly the configural unit for AB, so that a reasonably strong response will be observed. But when AB is presented, the stronger activation of the configural unit for AB than for either A or B will lead to the inhibitory link's exerting a more powerful influence on the output unit than do the excitatory links, and the CR will be weak.

In the next section, it will be shown how a configural theory of conditioning can explain a variety of experimental findings. The extent to which these findings can be explained by elemental theories of associative learning will also be considered, but first, a little more must be said about this type of theory. According to the Rescorla–Wagner (1972) theory, the associative strength of a CS will be the same no matter in what context it is presented. Hence, if a CS has entered into an excitatory association with a US, then however the CS is presented, it will have the same excitatory influence on the representation of the US. This influence can be masked by the presence of an inhibitor or augmented by the presence of another excitor, but the impact of the CS on the representation of the US is assumed to be the same in these and all other circumstances. To take account of the possibility that the associative properties of a CS do not remain constant if there is change in the stimuli that accompany it, several modifications to elemental theories of associative learning have been proposed. According to Hull's (1945) principle of afferent neural interaction, the presence of a stimulus can alter the manner in which a CS is perceived and thus can weaken the CR through generalization decrement. A similar proposal has been made more recently by Rescorla and Coldwell (1995), and a formal account of how the presence of one stimulus might affect the associative properties of another can be found in the theory of Wagner and Brandon (2001; see also Brandon & Wagner, 1998). This theory is a development of Wagner's (1976, 1981; Wagner & Rescorla, 1972) previous theorizing and extends considerably the range of experimental findings that can be explained from an elemental perspective.

In keeping with several other theories (e.g., Burke & Estes, 1957), Wagner and Brandon (2001) suggested that when a single stimulus is presented to an animal, it might excite a number of representational elements that can enter into excitatory or inhibitory associations. The changes in the strength of these associations are assumed to be determined by the Rescorla–Wagner (1972) equation, and responding in the presence of a stimulus or a compound is determined by the sum of the associative strengths of all the elements that are activated. When two stimuli are pre-

sent together, Wagner and Brandon proposed that their elements may interact in three different ways. They may all remain active and excite additional elements. This *added-element* model is equivalent to the Rescorla–Wagner model, with the additional elements serving the same role as unique or configural cues (Wagner & Rescorla, 1972). The second possibility is that when two stimuli are presented together, some of the elements in one stimulus will inhibit elements in the other stimulus. According to Wagner and Brandon, this possibility provides a computational account of Hull's (1945) principle of afferent neural interaction. They also show that this *inhibited-elements* model is formally very similar to the configural theory of Pearce (1994), so that both theories make similar predictions concerning the findings considered throughout this article. Finally, when two stimuli are presented together, some of the elements in one stimulus will be converted by the presence of the other stimulus into compound-unique elements. If rather few elements undergo this conversion, predictions from the *replaced-elements* theory will be similar to those from the Rescorla–Wagner theory, but if a substantial proportion of elements undergo this change, predictions will often be similar to those of configural theory. Wagner and Brandon believe that this replaced-elements model will prove to be more satisfactory than the other two, and for this reason, the following discussion will refer predominantly to this version of their theory.

## APPLICATION OF CONFIGURAL THEORY

### Overshadowing

Pavlov (1927) observed that conditioning with a CS may be less effective when it is accompanied by another CS for each conditioning trial than when conditioning is conducted with the CS by itself. From the little that has been said already, it should be evident that configural theory provides a straightforward explanation for this overshadowing effect. The discussion relating to Figure 2 indicated that after conditioning with a compound, AB, responding on trials with A by itself will be weaker than that on trials with the compound. Furthermore, the strength of the response to A will be weaker than it would have been if conditioning had been conducted with A by itself. A direct prediction from the foregoing analysis is that the effects of overshadowing will be observed after a single compound conditioning trial. Evidence that supports this prediction can be found in several experiments (e.g., James & Wagner, 1980; Mackintosh, 1971). Although this finding is incompatible with the explanation that is offered by the Rescorla–Wagner (1972) model for overshadowing (see Mackintosh, 1971), it has been explained from an elemental perspective by referring to the influence of attentional processes (James & Wagner, 1980; Mackintosh, 1971, 1975). One-trial overshadowing can also be explained by the replaced-elements theory of Wagner and Brandon (2001). Conditioning with a compound is predicted to have the result that only a portion of the elements activated by

each CS in isolation will enter into an association with the US. As a consequence, test trials with individual stimuli will result in a weaker CR after they have been presented for conditioning in compound rather than independently.

**Blocking**

In an experiment by Kamin (1969), an experimental group of rats received shock signaled by a single CS, A, before conditioning in which the compound AB signaled the same shock. This training was found to block or restrict conditioning with B, relative to a control group that was given compound conditioning without the prior training with A. To understand how blocking is explained by configural theory, consider first the treatment administered to the control group. For the purposes of discussion, it will be assumed that the salience of A and B is the same. Conditioning with AB in the control group will have the result that the configural unit excited by the compound will enter into an association of asymptotic strength,  $\lambda$ , with the US. The presentation of B by itself will then excite the AB configural unit to half its maximum value and will excite the US center to half the level that would occur in the presence of AB,  $0.5\lambda$ . For the experimental group, the initial training with A will have the result that the configural unit activated by this stimulus will enter into an asymptotic association with the US. When the compound AB is presented, then, according to the principles on which Equation 3 is based, the compound will excite the configural unit for A to half its maximum value and will result in a CR with a strength of  $0.5\lambda$ . The trials with AB will show an increase in the associative strength of the configural unit for AB, which will cease when it has reached  $0.5\lambda$ . Presenting B by itself will then activate the configural unit for AB to half its maximum value and will result in the US center's being activated to  $0.25\lambda$ , which is half the level predicted for the control group.

Configural theory also predicts that blocking with B will be observed if the trials with A+ and AB+ are intermixed within the same session. Evidence in support of this prediction can be found in Wagner (1969). Interestingly, the theory further predicts that blocking will not be found with this methodology if it is modified in a relatively simple way. Consider an experiment by Rescorla (1989), who gave pigeons intermixed trials with A+ and AB+ after they had received A+ B+ C+ training (see Table 1, upper row). Test trials revealed that responding to B was similar to that to C, which indicates that the blocking treatment in Stage 2 did not attenuate at all the associative properties acquired by B in the initial stage of the experiment. When

AB is presented at the outset of Stage 2, configural theory predicts that A and B will each contribute half of their associative strengths,  $0.5\lambda$ , to the compound, so that it will elicit a CR of asymptotic strength when it is first presented. The configural unit for AB in these circumstances will gain no associative strength, and the associative properties of B will, therefore, be unaffected by compound conditioning.

In a study that is similar to the one reported by Rescorla (1989), Pearce and Redhead (1995) first trained rats with a B+ AB- C+ discrimination, prior to giving them intermixed trials of the form AB+ A+ C+ (see Table 1, lower row). For this study, B and C were auditory stimuli, and A was a visual stimulus. Responding during subsequent test trials was stronger in the presence of B than in the presence of C. In terms of configural theory, there will be scope for AB to acquire associative strength during the second stage of the experiment. When B is then presented for testing, it will activate the configural unit for B fully and will excite an asymptotic CR by virtue of the initial training with B. In addition, the presentation of B will arouse, to some extent, the excitation associated with the configural unit for AB. The combination of these two sources of excitation can then be expected to produce a CR of super-normal strength.

The results from the two experiments, whose design is summarized in Table 1, are not consistent with predictions that can be derived from the Rescorla-Wagner (1972) theory. During Stage 2 of each experiment, the conditioning trials with A would be predicted to weaken substantially the associative properties previously acquired by B, rather than leaving them unaffected or strengthened. In contrast, the replaced-elements theory of Wagner and Brandon (2001) can, in principle, explain the results from the two experiments. Assume that during the AB+ trials in Stage 2, half of the elements of each CS are replaced. For the experiment by Rescorla (1989), the overall associative strength of AB will then be at the asymptotic value set by the US, and there will be no change in the associative strength of the elements activated by B. When B is then presented for testing, all of its elements will again be activated, and because none of them will have lost any associative strength, the magnitude of the CR should be similar to that observed at Stage 1. The explanation for the results from the experiment by Pearce and Redhead (1995) is essentially the same as that just outlined, except that at the outset of Stage 2, the low overall associative strength of AB will allow some of the B elements that are activated to gain associative strength. The increase in the associative proper-

**Table 1**  
**The Designs of Two Experiments in Which the Effects of**  
**Intermixing Trials of the Form AB+ and A+ on Subsequent**  
**Responding With B Are Investigated**

Study	Stage 1	Stage 2	Test
Rescorla (1989)	A+ B+ C+	AB+ A+	B and C
Pearce & Redhead (1995)	B+ AB- C+	AB+ A+ C+	B and C

ties of the elements of B that are activated during AB trials, together with the fact that the associative properties of some of the elements of B will be unaffected during Stage 2, because they will be inhibited, will then ensure that an abnormally strong response is recorded during the test trials with B.

It should be apparent that these explanations depend critically on the extent to which stimulus A can be assumed to inhibit elements in B; the results of the experiment by Rescorla (1989) can be explained only if the presence of A is able to inhibit half of the elements of B. The problem posed by this conclusion is that, for their replaced-elements theory, Wagner and Brandon (2001) assumed that for every different stimulus paired with stimulus A, a different set of elements will be inhibited. Once it has been granted that stimulus B inhibits half the elements of A, it becomes apparent that the theory can be applied only to experimental designs in which A is paired with no more than two stimuli, which might be seen as an unreasonable restriction.

The results from Pearce and Redhead (1995) raise a further issue concerning the theoretical proposals of Wagner and Brandon (2001). The replaced-elements theory is able to explain the outcome of this experiment, provided that the proportion of elements that one stimulus inhibits in another is relatively high. One factor that has been said to influence the degree to which one stimulus will inhibit the elements of another stimulus is whether they can be regarded as separable or integral components of the compound to which they belong (Myers, Vogel, Shin, & Wagner, 2001). Separable components of a compound are assumed to inhibit fewer of each other's elements than do integral components. The following distinction is made between the two types of stimuli: "Separable components are ones that appear to be processed the same in compound as in isolation. Integral components are ones that appear to combine in a more holistic fashion, and not to be processed the same in compound as in isolation" (Myers et al., 2001, p. 43).

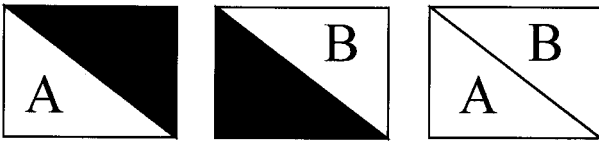
On this basis, Myers et al. (2001) concluded that the elements of one stimulus are more likely to be inhibited by another if they belong to the same, rather than to different, modalities. For the experiment by Pearce and Redhead (1993), A was a visual stimulus, and B was an auditory stimulus. Given that the compound is composed of stimuli from different modalities, the most reasonable conclusion to draw is that it is composed of separable components and that there will be rather little interaction between their elements. According to the replaced-elements theory of Wagner and Brandon (2001), therefore, the outcome of the study by Pearce and Redhead (1995) should have been more in keeping with predictions that can be derived from the Rescorla-Wagner (1972) theory than from configural theory.

### Summation

A potentially serious problem for configural theory is posed by the results from studies of summation, which is

said to occur if responding for a compound of two stimuli that have separately been paired with a US is stronger than that for either stimulus alone. Provided that the stimuli are of equal salience, it follows from configural theory that half of the associative strength of each stimulus will generalize to the compound, and the combination of these strengths will then result in a response that is no greater than that seen for the components. There are numerous successful demonstrations of summation (e.g., Kehoe, 1982, 1986; Konorski, 1948, pp. 109–114; Whitlow & Wagner, 1972), which poses a challenge to the account offered by configural theory of the way in which responding for a compound CS is determined. A further problem posed by studies of summation concerns the finding that summation is often more pronounced when the compound is composed of stimuli from different modalities than when composed of stimuli from the same modality (Kehoe, Horne, Horne, & Macrae, 1994). In contrast to this result, the most direct prediction from configural theory is that responding during a compound should be somewhat stronger if its components are from the same modality, rather than from different modalities. To understand this prediction, assume that stimuli from the same modality share a common feature, C. The patterns of stimulation present during conditioning can thus be conceptualized as AC and BC and, at testing, as ABC. If all three elements are of similar salience, then on the basis of the equations presented earlier, the strength of the CR predicted for a test trial with ABC is  $1.07\lambda$ .

Despite these problems, there are a number of results that demonstrate that configural theory can be used with considerable success to predict the outcome of more complex studies of summation than those just mentioned. Many of these experiments have been conducted using autoshaping with pigeons, for which it is generally found that a compound of two stimuli presented on a display behind a response key yields a similar rate of responding to that with either stimulus alone (e.g., Aydin & Pearce, 1994, 1995, 1997; Rescorla & Coldwell, 1995). Of course, this outcome can be readily explained by configural theory. Further support for the account provided by configural theory for the way in which responding during a compound is determined can be found in an experiment by Pearce, Aydin, and Redhead (1997). In this study, pigeons received training in which the delivery of food was signaled by pairs of stimuli (AB, AC, and BC), as opposed to individual stimuli. Subsequent test trials revealed stronger responding with ABC than with a control compound composed of three stimuli, DEF, that had consistently been paired with food. Because each training compound provided two thirds of the elements that were present in the test compound, configural theory would predict that there would be considerable generalization to ABC from the pairs of stimuli, with the result that a response of greater than asymptotic magnitude should be observed. Indeed, a computer simulation described by Pearce et al. (1997) revealed that the associative strength of ABC, when it is first presented, is predicted by the theory to be equal to  $1.33\lambda$ .



**Figure 3.** The three different patterns that were presented on a television screen during conditioning trials with A (left-hand panel), and B (center panel) and test trials with AB (right-hand panel) in the experiment by Aydin and Pearce (1997).

A rather different demonstration of summation, using both autoshaping with pigeons and appetitive conditioning with rats, was provided by Pearce, Adam, Wilson, and Darby (1992; see also Pearce et al., 1997). Subjects first received an AB+ BC+ C− discrimination prior to test trials with ABC, which revealed a stronger response than with a control compound, DEF, that, for example, had been consistently paired with food. Once again, a strong response is predicted by configural theory to be observed during testing with ABC, because the similarity of the training compounds to the test compound will ensure substantial generalization to the test compound. A more detailed account of the rationale behind this prediction can be found in Pearce et al. (1992), who show that the associative strength predicted for ABC, when it is first presented, is 1.33λ.

According to configural theory, an AC+ BC+ C− discrimination will result in strong responding during ABC only if the salience of C is approximately the same as that of A and B. As the salience of C declines, the strength of responding during ABC is predicted to diminish. In the limiting case, in which C has no salience, training will be of the form A+ B+, and as was noted above, the theory predicts that responding for AB will be no different from that for A or B alone. To test this analysis, Darby and Pearce (1995, 1997) trained rats with an AC+ BC+ C− discrimination, using appetitive conditioning. For some subjects, the duration of C was the same as that for A and B. This manipulation was intended to ensure that the salience of C was relatively high, and test results revealed strong responding during ABC. For other subjects, steps were taken to reduce the salience of C by presenting it throughout the entire experimental session (Darby & Pearce, 1995) or by repeatedly presenting it before the start of discrimination training (Darby & Pearce, 1997). Both manipulations resulted in relatively weak responding during ABC.

Typically, studies of summation that use autoshaping with pigeons involve training in which brightly colored stimuli are presented against a dark background on a television monitor, which is also dark during the intertrial interval. By way of example, Figure 3 shows the stimulation provided by a television screen in a summation study by Aydin and Pearce (1997) on conditioning trials with Stimulus A (left-hand panel), and Stimulus B (center panel) and on the test trials with AB (right-hand panel). If the entire pattern of stimulation on the screen is regarded as the

configuration that signals food, the presence of the dark screen during training ensured that the patterns created on trials with A and B were quite different from the pattern created by the test trials with AB: For the training trials, the screen was half bright and half dark, whereas for testing, it was entirely bright. According to configural theory, therefore, generalization from the training patterns to the test patterns will be restricted, and summation should not occur. One prediction that follows from this analysis is that if the region of the screen that is not filled by the experimental stimuli is bright, rather than dark, both during each trial and during the intertrial interval, the similarity of the training patterns to the test pattern will be increased and will permit summation to occur. This prediction has been confirmed by Aydin and Pearce (1997), Pearce, George, Redhead, Aydin, and Wynne (1999), and Pearce, Redhead, and George (2002), who show formally how this result can be explained by configural theory. To provide a further test of the account offered by configural theory for summation, Pearce, Redhead, and George repeated the experiment that has just been described, except that A and B were dark, rather than bright. According to the foregoing account, if the television screen is bright in the regions not occupied by the experimental stimuli, there will be little generalization to the dark test compound, and summation will not be seen. On the other hand, if the television screen is dark in the regions not occupied by the stimuli, the patterns created during training will be similar to the dark test compound, there will be substantial generalization to the test compound, and summation will be seen. Both of these predictions were confirmed.

Configural theory is thus able to explain a variety of summation results, and it is worth considering whether it is possible to use this theory to explain the more conventional demonstrations of summation considered at the beginning of this section. One explanation can be developed by referring to the experiment by Pearce et al. (1992), in which training with an AC+ BC+ C− discrimination resulted in summation on test trials with ABC. In essence, this experiment can be regarded as being equivalent to a conventional summation experiment in which A and B are the experimental stimuli and C is the context in which they are presented. Configural theory was shown earlier to predict an abnormally strong response to ABC if the salience of C is high, but not if it is low. Thus, by taking account of the experimental context, configural theory is able to predict summation after conditioning with two stimuli, but only if it can be assumed that the context is salient. It is also of interest to note that once conditioning with two stimuli is conceptualized as an AC+ BC+ C− discrimination, responding on test trials with ABC is predicted to be stronger when A and B belong to different modalities than when they belong to the same modality. This prediction, which is opposite to one made earlier concerning the effects of conditioning with A and B and testing with AB, stems from two implications of configural theory: First, the AC+ BC+ C− discrimination will be easier to solve if

there is some generalization between A and B, and second, responding during ABC will be weaker if the AC+ BC+ C- discrimination is easy than if it is difficult.

To what extent is it reasonable, therefore, to assume that the context is salient during training prior to a test for summation? Such an assumption may be reasonable for studies involving instrumental conditioning (Hamm, Porter, & Oster, 1978; Long & Allen, 1974; Meltzer & Hamm, 1976; Millier & Beale, 1977; Strub & Barske, 1977; Weiss, 1977; Wiltz, Boren Moerschbaeher, Creed, & Schrot, 1973). Consider, for example, a study in which rats are first trained to press a lever for food during periods when A or B are present, but not when they are absent. The lever can be regarded as a contextual cue, and since it must be pressed if food is to be delivered, it is likely that the subjects will pay considerable attention to this stimulus. Thus, for operant studies of summation, at least a component of the contextual stimuli may be of considerable salience, which would then allow configural theory to explain summation in these circumstances. Conceivably, a similar explanation could account for successful demonstrations of summation in which Pavlovian conditioning is used, such as conditioned suppression, where training is superimposed on an instrumental baseline (Hendry, 1982; Wagner, 1971). But this type of explanation becomes more strained for demonstrations of summation using Pavlovian conditioning that does not take place against an instrumental baseline. Examples of these studies include eyelid conditioning with rabbits (Kehoe, 1982; Kehoe et al., 1994; Whitlow & Wagner, 1972), salivary conditioning with dogs (Konorski, 1948, pp. 109–114), and appetitive Pavlovian conditioning with rats (Rescorla, 1997). In these experiments, the target stimuli are presented against a background that is present throughout each experimental session, and it is likely that the salience of the background will be low and summation will not be seen. If configural theory is to explain these demonstrations of summation, it would have to find a special reason for assuming that the salience of the background cues is high. For instance, the fact that rabbits must be restrained throughout conditioning might provide justification for this assumption, as far as demonstrations of summation with this species is concerned. Alternatively, certain demonstrations of summation could occur for nonassociative reasons. Responding for a compound might be stronger than that for its components because the strength of the CR is related to the intensity of the CS (Hull, 1949), and the intensity of a compound will necessarily be greater than that of its constituents. Another possibility is that inhibition of delay might serve to weaken responding during conditioning with a single CS, so that when it is presented with another CS, any evidence of summation might be a consequence of disinhibition of inhibition of delay (Aydin & Pearce, 1995). Whether there is any merit in these and other possible explanations for summation remains to be seen. For the present, these suggestions are offered principally as a means of showing that successful demonstra-

tions of summation may not lie beyond the scope of configural theory.

Elemental theories provide a straightforward explanation for summation because they assume that the associative strengths of two CSs will combine when they are presented together. Furthermore, given the assumption that the elements of two or more stimuli are more likely to inhibit each other when the stimuli belong to the same modality, rather than to different modalities, the replaced-elements theory of Wagner and Brandon (2001) is able to explain most of the successful and unsuccessful demonstrations of summation reviewed above. One problem that confronts this theory derives from the finding that summation in autoshaping is influenced by the similarity of the training stimuli to the background on the television screen on which they are presented. If the training stimuli and the background are both dark or both bright, summation can be found, but if one of them is dark and the other bright, summation is unlikely to take place (Pearce, Redhead, & George, 2002). For this pattern of results to be explained by the replaced-elements theory, the training and the background stimuli must be assumed to inhibit fewer of each other's elements when they are similar than when they are different. It is not at all clear that this assumption can be justified. A further problem for this theory is posed by published and unpublished experiments in my laboratory that have tested for appetitive summation by using Pavlovian conditioning with rats and magazine activity as the CR (e.g., Pearce, George, & Aydin, 2002). If conditioning is conducted with a single visual and a single auditory CS, summation is not observed when these stimuli are presented together.<sup>1</sup> Because the stimuli are from different modalities, the theory of Wagner and Brandon (2001) predicts that summation should be seen in these conditions. The conflict between these findings and the successful demonstrations of summation using, say, rabbit eye-blink conditioning suggests that the factors that determine whether or not summation will be observed remain to be fully understood. Until this understanding is achieved, it may be premature to conclude that either a configural or an elemental theory provides the correct account for summation.

## Discrimination Learning

### Conditioned inhibition and retroactive interference.

When animals are presented with a discrimination of the form A+ AB-, responding is eventually considerably stronger for A than for AB (Pavlov, 1927). Furthermore, if B should be presented by itself, it will pass both a summation and a retardation test for conditioned inhibition (Rescorla, 1969). Training with an A+ AB- discrimination will, in terms of configural theory, result initially in A's acquiring excitatory properties. Subsequent trials with AB will then result in its arousing a measure of excitation through generalization from AB, and to counter this excitation, AB will enter into an inhibitory association. The discrimination will be solved when A has sufficient exci-



tatory strength to elicit a response of asymptotic magnitude and AB has sufficient inhibitory strength to counter completely the excitation that generalizes to it from A. If B is presented alone, it will arouse inhibition through generalization from AB and will effectively possess negative associative strength. The solution of an A+ AB- discrimination, in terms of the Rescorla–Wagner (1972) theory, is said to depend on A’s acquiring positive associative strength and B’s acquiring an equivalent magnitude of negative associative strength.

Pearce and Wilson (1991; see also Wilson & Pearce, 1992) described a series of appetitive conditioning experiments with rats and pigeons whose design would allow one to choose between these two accounts of the way in which an A+ AB- discrimination is solved. In one experiment, subjects first received an A+ AB- discrimination. They then received excitatory conditioning with B before being returned to the original discrimination. According to the Rescorla–Wagner (1972) theory, the second stage of training would erase completely the negative associative strength originally acquired by B and would replace it with positive associative strength. When AB is presented for the final stage of the experiment, the associative strengths of A and B would be predicted to combine, resulting in stronger responding than that for A by itself. That is, conditioning with B in the second stage of the experiment would be predicted not only to abolish the effects of the original discrimination, but also to reverse them. Contrary to this prediction, when the discrimination was introduced for the test phase of the experiment, there was some evidence of a disruptive influence of the training with B by itself, but there was certainly no hint that this treatment resulted in stronger responding for AB than for A. Indeed, relative to a variety of control conditions, the results from the final stage revealed that there were considerable savings from the effects of the first stage of training. Conditioning with B disrupted, but did not abolish or reverse, the effects of the original exposure to the A+ AB- discrimination. This pattern of results can be readily explained by configural theory. The trials with B during the second stage will permit it to enter into an excitatory association. When AB is presented for the final stage of the experiment, its inhibitory properties will again counter the excitation that generalizes from A, but

there will be no residual inhibition left to counter the excitation that will also generalize to AB as a result of the training with B. As a consequence, responding for AB will be stronger than at the end of the first stage, but it will be weaker than that for A, because the generalization of excitation from B to AB will be incomplete.

As was noted above, the replaced-elements theory of Wagner and Brandon (2001) may make predictions similar to those of configural theory (Pearce, 1987, 1994), providing that two stimuli together inhibit a good proportion of each other’s elements. Therefore, the results from the experiments just described can, in principle, be explained by this theory. In some of the experiments, however, A and B were from different modalities, and it is not clear that an explanation for the results of these experiments in terms of the replaced-elements theory is justified. More specifically, if A and B can be regarded as being separable components of the compound AB, then the replaced-elements theory is led toward predicting an outcome of these experiments that is similar to the prediction made by the Rescorla–Wagner (1972) theory, rather than to that made by configural theory.

The problem posed by the experiments of Pearce and Wilson (1991) for the Rescorla–Wagner (1972) theory is an instance of a more general problem that applies to a variety of theories of learning. McCloskey and Cohen (1989) have argued that connectionist networks based on the back-propagation algorithm (e.g., Rumelhart, Hinton, & Williams, 1988) predict unrealistically high amounts of retroactive interference when training episodes involving similar stimuli are presented successively. The Rescorla–Wagner (1972) model is formally equivalent to these networks (see Gluck & Bower, 1988; Maki & Abunawass, 1991; Sutton & Barto, 1981) and is thus prone to making the same erroneous prediction. In contrast, as Page (2000) points out, any theory that assumes that associative learning depends on a configural—or as he refers to it, “localist”—representation’s entering into a single association with an outcome is able to provide a more accurate account of the effects of retroactive interference.

**Similarity and discrimination learning.** One factor that, at least intuitively, would seem to be critically important for determining the ease with which a discrimination will be solved is the similarity of the signals for reward

**Table 2**  
**Four Experimental Designs That Have Been Used to Study the Relationship**  
**Between the Similarity of the Signals for Reward and Nonreward and the**  
**Ease of a Discrimination Between Them**

Study	Design
Pearce and Redhead (1993), Experiment 1	
Group E	A+ AB-
Group C	AC+ ABC-
Pearce and Redhead (1993), Experiment 3	
Group E	A+ B+ AB-
Group C	AC+ BC+ ABC-
Redhead and Pearce (1995a), Experiment 1	A+ BC+ ABC-
Redhead and Pearce (1995a), Experiment 2	A+ B+ C+ AB+ AC+ BC+ ABC-

and nonreward. Discrimination learning is likely to progress more rapidly when the signals for reward are similar than when they are different. This relationship can be readily predicted by configural theory. If two patterns are similar, there will be more generalization of excitation and inhibition between them than if they are different, and the greater the degree of this generalization, the more slowly will the discrimination develop. By way of contrast, Pearce (1994) has argued that there are certain occasions when the Rescorla–Wagner (1972) model makes counterintuitive predictions about the relationship between similarity and the ease with which a discrimination will be mastered.

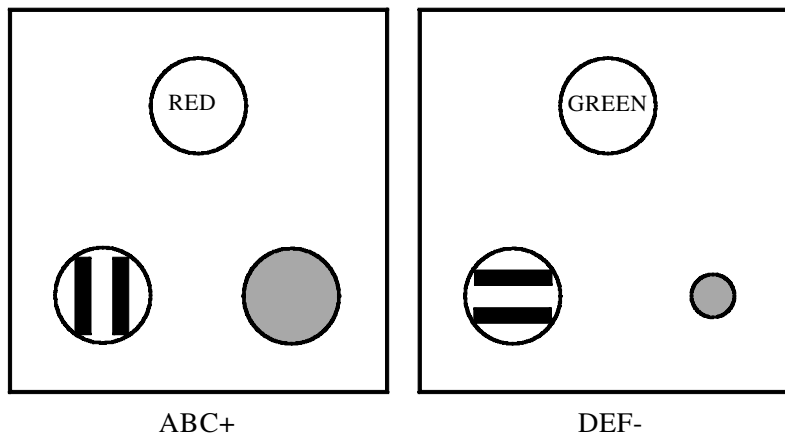
Examples of some of the designs that pose a problem for the Rescorla–Wagner (1972) model are summarized in Table 2. The top design outlines two discriminations that were given to two groups of pigeons. Group E received a straightforward feature-negative discrimination with food presented after A, but not after AB. Group C was trained with the same discrimination, except that a common element, C, was added for both trials. According to the relationship expressed in Equation 4, the similarity between A and AB is 0.5, whereas between AC and ABC, it is 0.66. Thus, configural theory predicts that the discrimination given to Group E will be acquired more readily than the one given to Group C. The results from the experiment confirmed this prediction. In contrast to these results, the Rescorla–Wagner (1972) theory predicts that the discrimination will be easier to solve for Group C than for Group E.

The theories also make opposing predictions concerning the discrimination summarized in the second design from the top of Table 2. The negative-patterning discrimination given to Group E is predicted by configural theory to be easier than the one given to Group C, whereas the Rescorla–Wagner (1972) theory predicts the opposite outcome. In fact, the discrimination was acquired more readily by Group E than by Group C. For both of these experimental designs, the discriminative stimuli were arrays of randomly located colored dots presented on a television screen behind a response key. Thus, A might have been an array of red dots, AB a mixture of red and green dots, and ABC a mixture of red, green, and white dots. Conceivably, the presence of C for the trials with Group C made it more difficult to identify the presence and absence of A and B, which would then account for the poor acquisition of the discrimination by this group in each experiment (Myers et al., 2001; Rescorla & Coldwell, 1995). This type of explanation is unlikely, however, to account for the outcome of an experiment by Rescorla (1972), which was of similar design to that used by Pearce and Redhead (1993, Experiment 3), except that A and B were auditory stimuli and C was a light. In keeping with the results reported by Redhead and Pearce (1993), Rescorla (1972) found that Group C acquired the discrimination more slowly than did Group E, but on this occasion, it is difficult to argue that the presence of the visual stimulus, C, made it hard to detect whether or not the auditory stimuli, A and B, were present.

The remaining designs in Table 2 summarize discriminations that were given to single groups of pigeons. For the A+ BC+ ABC– discrimination, configural theory predicts that the discrimination between A and ABC will be mastered more readily than that between BC and ABC, whereas the Rescorla–Wagner (1972) theory predicts the opposite outcome. The results confirmed the prediction from configural theory. In a sense, the experiment that has just been considered is a negative-patterning discrimination in which one signal for the US, A, is less salient than the other signal, BC. It should not be surprising to find, therefore, that according to configural theory, if a negative-patterning discrimination is conducted with stimuli of different salience, the discrimination between the weaker stimulus and the nonreinforced compound will be acquired more readily than the discrimination between the stronger stimulus and the compound. This prediction has been confirmed in an experiment by Redhead and Pearce (1995b), for which the salient stimulus was a tone and the weak stimulus was created by turning off the houselight (see also Delamater, Sosa, & Katz, 1999). The outcome of the experiment by Redhead and Pearce (1995b) is of interest because the use of stimuli from different modalities makes it difficult to explain the results by assuming that they were a consequence of one stimulus' interacting perceptually with another. It is also difficult to explain this finding by the replaced-elements theory of Wagner and Brandon (2001).

Turning now to the discrimination summarized in the bottom row of Table 2, configural theory predicts that the discrimination between the individual stimuli and ABC will progress more readily than the discrimination between the pairs of stimuli and ABC, and yet again, the Rescorla–Wagner theory predicts the opposite outcome. The results were consistent with the predictions from configural theory by showing that as the discrimination was acquired, responding was relatively slow for ABC, somewhat faster for the pairs of stimuli, and fastest for A, B, or C.

The stimuli used for the experiment by Redhead and Pearce (1995a) were red, green, and white dots intermixed among each other, but the pattern of results that was reported does not depend on this particular set of stimuli. Pearce and George (in press) conducted an autoshaping experiment based on the design shown in the bottom panel of Table 2, but additional stimuli were used to create the following trials: D– E– F– DE– DF– EF– DEF+. The stimuli were shown on a television screen behind a response key. The left-hand panel of Figure 4 depicts the stimuli presented for reinforced trials with ABC, and the right-hand panel shows the stimuli presented for nonreinforced trials with DEF.<sup>2</sup> Trials with pairs of stimuli and individual stimuli involved removing the unwanted stimuli from the screen. The results from this study were essentially the same as those for the experiment by Redhead and Pearce (1995a) and thus support the predictions from configural theory. Furthermore, whereas the results reported by Redhead and Pearce (1995a) were obtained with stim-



**Figure 4.** The stimuli that were used by Pearce and George (in press) for a discrimination involving pigeons.

uli from the same dimension, in the study by Pearce and George, they were from different dimensions: size, orientation, and color.

The experimental design shown in the bottom row of Table 2 does not always produce results that are consistent with predictions that can be derived from configural theory. Myers et al. (2001) have also conducted an experiment based on this design, using eye blink conditioning with rabbits, and their results were more in keeping with predictions that can be derived from the Rescorla–Wagner (1972) theory than with those from configural theory. That is, during the acquisition of the discrimination, responding was stronger on trials with the pairs of stimuli than on those with the individual stimuli. To explain the contrast between these results and those reported by Redhead and Pearce (1995a), Myers et al. referred to the replaced-elements theory of Wagner and Brandon (2001). In the study with rabbits, the stimuli were from three different modalities, and the prediction made by the replaced-elements theory for this experiment was said to be similar to that made by the Rescorla–Wagner theory. On the other hand, the three stimuli used in the experiment by Redhead and Pearce (1995a) were from the same modality, and Myers et al. concluded that the replaced-elements theory would make similar predictions about the experiment as configural theory. Unfortunately, Myers et al. did not report the outcome of any computer simulations to support their argument, and it remains to be demonstrated that replaced-elements theory is able to explain the findings by Redhead and Pearce (1995a) and George and Pearce (2002).<sup>3</sup>

Further experiments are required before unequivocal theoretical conclusions based on the designs shown in Table 2 can be drawn. For the present, experiments using stimuli from different modalities or from different dimensions suggest that the failures to confirm predictions of the Rescorla–Wagner (1972) theory concerning the designs shown in Table 2 were not due to an interaction between

individual stimuli, where the interaction either was perceptual or consisted of one stimulus' inhibiting the representational elements of the other. Of course, this conclusion rests on the assumption that there will be rather little interaction between stimuli that belong to different dimensions or different modalities. Such an assumption is in keeping with the arguments developed by Rescorla and Coldwell (1995) and Myers et al. (2001), but it might be argued that it is too restrictive. Perhaps the presence of one stimulus will influence quite markedly the characteristics of any other stimulus that accompanies it, no matter whether they belong to the same or different dimensions or modalities. For this argument to be taken seriously, there is a need to specify precisely the conditions under which one stimulus will interact, and to what extent, with another.

The conflicting results concerning the discrimination shown in the bottom row of Table 2 are particularly troubling. One factor that might exert an important influence on the course of the acquisition of this discrimination is whether summation can be readily found with stimuli when they are presented together after having been separately paired with a US. In the experiment by Myers et al. (2001), the subjects were rabbits, and the stimuli were from different modalities. A number of studies have provided clear evidence of summation, using eye blink conditioning with rabbits and stimuli from different modalities, and it should perhaps not be surprising to find that responding for the pairs of stimuli was stronger than that for the individual stimuli in the study by Myers et al. In contrast, summation is rarely found with visual stimuli that are presented on a display behind a response key for autoshaping with pigeons. In view of this outcome, it is perhaps not surprising that Redhead and Pearce (1995a) and Pearce and George (in press) failed to find stronger responding for the pairs of stimuli than for the individual stimuli in their experiments based on the design summa-

rized in the lower panel of Table 2. The outcome of this discrimination may thus depend on whether or not the training conditions promote summation, and as was noted earlier, these conditions are not yet fully understood.

### Summation Following Discrimination Training

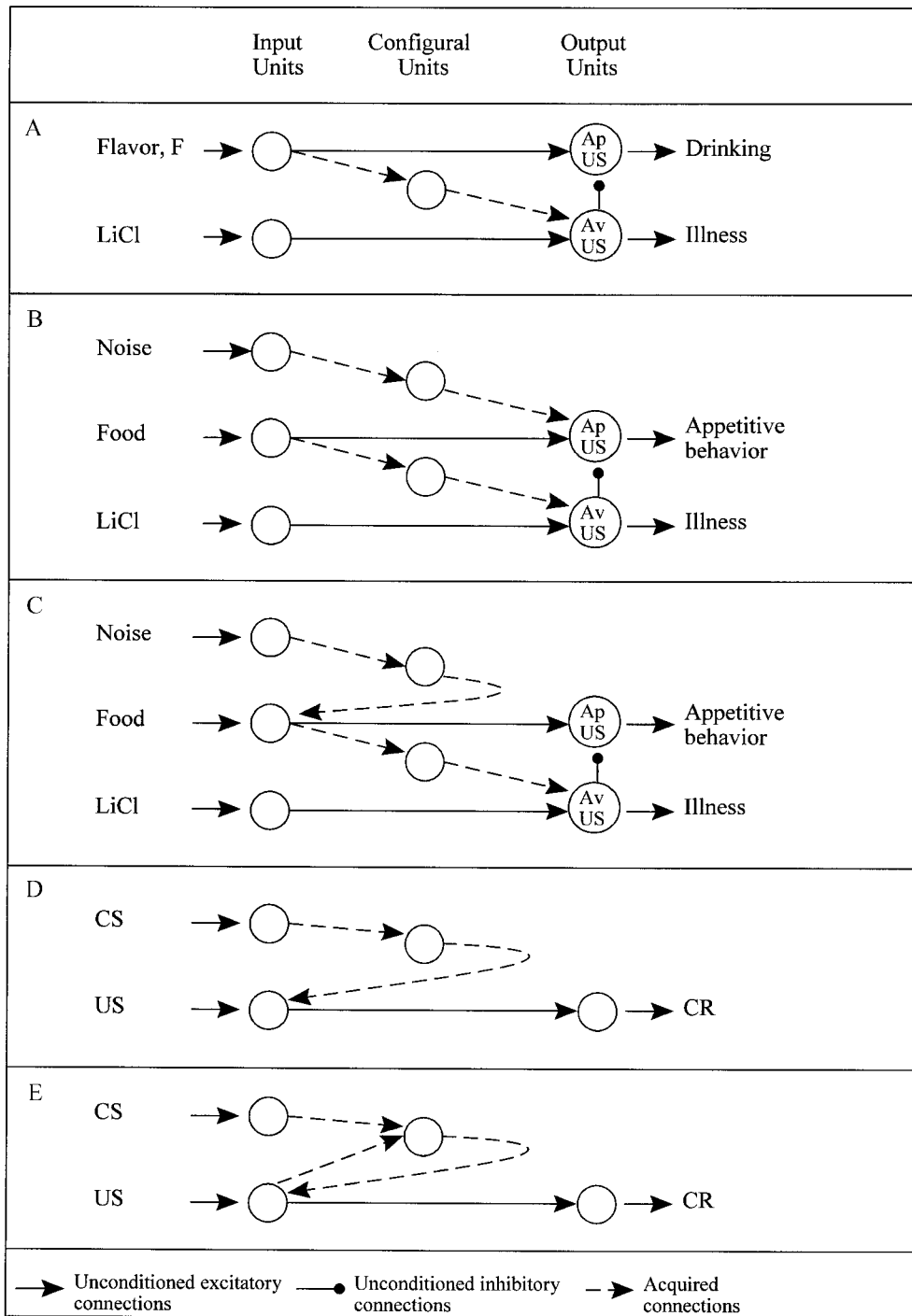
In a number of studies, animals have been trained with a discrimination involving compounds of two or more stimuli before being given test trials with a new combination of stimuli. According to the Rescorla–Wagner (1972) theory, responding during the test compound will be determined by the sum of the associative properties of its components, whereas configural theory predicts that it will be determined by generalization from the patterns of stimulation presented for the discrimination. Occasionally, the two theories make opposing predictions about the outcome of the test trials.

Pearce and Wilson (1990a) first trained pigeons with an  $A+ AB- BC+$  discrimination, using autoshaping, before test trials were conducted in which responding for BC was compared with that for C. They found that responding for BC was stronger than that for C. From the point of view of configural theory, this finding is not surprising, because responding for C will be determined entirely by generalization from the associative properties of BC. The absence of B will then lead to C by itself exciting a weaker response than BC. Using a related discrimination,  $A+ AB- ABC+$ , Nakajima (1997, Experiment 3) and Nakajima and Urushihara (1999, Experiments 4a and 5a) found that responding on test trials was faster for BC than for C. In the experiments reported by Nakajima and Urushihara, elements belonging to two modalities were used, and the effect just described was found when A and B were from the same modality and C was from the other modality. However, in other experiments, Nakajima and Urushihara trained animals when A belonged to one modality and B and C belonged to another modality. In these experiments, the CR was stronger in the presence of C than in the presence of BC. Nakajima and Ushihara explained the overall pattern of their results in terms of configural theory, by making the reasonable assumption that there would be generalization between stimuli belonging to the same modality. In the task with A and B from the same modality, responding on the test trials with BC was predicted to be strong, because there would be excitatory generalization to this compound from both A (because of the similarity between A and B) and ABC. Furthermore, responding for C was predicted to be relatively weak, because only a relatively small proportion of the excitation associated with ABC would generalize to C. On the other hand, when B and C were from the same modality, during training there would be substantial generalization of excitation from ABC to AB, which would require AB to acquire a substantial level of inhibition if responding for this compound is to be weak. The generalization of inhibition from AB to BC, together with the lack of excitation generalizing from A to BC, would then ensure that responding would be weaker for BC than for C.

The findings that have just been mentioned are difficult to explain with elemental theories of learning. The results reported by Pearce and Wilson (1990a) and by Nakajima (1997), in which responding was stronger on trials with BC than on those with C, are difficult to explain with the Rescorla–Wagner (1972) theory, because it predicts that the training will convert B into a conditioned inhibitor, so that it should suppress, rather than augment, responding during C. The replaced-elements theory of Wagner and Brandon (2001) is able to predict the results reported by Pearce and Wilson (1990a), providing that the proportion of elements that are altered when two stimuli are combined is not too low. But the theory is unable to account for the effects of manipulating the modality of the stimuli in the experiments of Nakajima and Urushihara (1999). Computer simulations of the replaced-elements theory consistently have predicted that responding during test trials with BC will be stronger than responding during those with C, even when generalization is allowed between pairs of stimuli (by assuming that they excite a common element) and even when it is assumed that stimuli from the same modality modify more of each other's elements than do stimuli from different modalities.

Finally, Pearce and Wilson (1990b) first trained pigeons with an  $ABC+ A+ B- C-$  discrimination, before presenting A, AB, BC, and C for testing. The important findings from the test trials were that responding to A was stronger than that to AB and that responding to C was weaker than that to BC. Thus, B suppressed responding to A and enhanced responding to C, which was anticipated by configural theory. In brief, responding to AB was expected to be weak, because of the generalization of the effects of the nonreinforced trials with B. On the other hand, responding to BC was predicted to be stronger than that to C, because the closer similarity of BC to ABC than of C to ABC would permit a greater generalization of excitatory strength from ABC to BC than from ABC to C. One reason for mentioning this study is that the findings again pose a problem for the Rescorla–Wagner (1972) theory. The faster responding to BC than to C implies, according to this theory, that B has excitatory properties, which then makes it difficult to explain why this stimulus had the effect of suppressing responding to A. This pattern of results can, however, be explained by the replaced-elements theory of Wagner and Brandon (2001).

The configural theory of Pearce (1987, 1994) was formulated originally to account for the effects of stimulus generalization in Pavlovian conditioning—in particular, for the way in which the properties of a CS would be affected by a change in the stimuli that accompanied it—and for the way in which discriminations are solved. Experimental tests of the theory have tended to focus on the predictions it makes that are different from those made by the Rescorla–Wagner (1972) theory. In developing these tests, a number of new experimental paradigms have been employed. On the whole, the novel predictions derived from configural theory concerning these tests have been confirmed, which suggests that the theory provides a use-



**Figure 5.** Possible modifications to the configural network proposed by Pearce (1994). (A) The connections that will form during taste aversion conditioning if unconditioned stimuli (USs), are allowed to excite input units. (B) The connections that will form in the reinforcer devaluation study by Holland and Straub (1979) if configural units always become connected to output units. (C) The connections that will form in the study by Holland and Straub if configural units can become connected to input units. (D) The connections that will form whenever a conditioned stimulus (CS) and a US are paired if configural units can become connected to input units. (E) The connections that will form whenever a CS and a US are paired if the activation of the input unit for the US, before the US is presented, is allowed to activate a configural unit.

ful account of the mechanisms of generalization and discrimination for Pavlovian conditioning. The purpose of the next section is to consider a number of ways in which the connectionist network shown in Figure 2 can be developed in order to allow it to be applied to a broader range of phenomena.

## POSSIBLE MODIFICATIONS TO THE NETWORK

### Unconditioned Stimuli Activate Input and Output Units

For the network shown in Figure 2, neutral stimuli, such as a CS, are assumed to activate input units, and biologically significant events, such as a US, are assumed to excite output units. As it stands, therefore, the network would be unable to account for successful conditioning in which a biologically significant event is used as a CS for Pavlovian conditioning. Demonstrations of taste aversion conditioning (Garcia & Koelling, 1966), to cite one example, show forcefully that conditioning can be successful in these circumstances, and the network shown in Figure 2 needs to be modified to take account of such findings.

An obvious modification is to allow all stimuli to excite units at the input layer of the network, but some of these will have unconditioned links with output units. Suppose a thirsty rat is allowed to drink water with a distinctive flavor, *F*, before receiving an injection of a mild toxin, lithium chloride (LiCl). The solution is a US that will excite an output unit that is responsible for drinking, and LiCl is a US that will excite an output unit that is responsible for exciting a variety of responses related to illness, including cessation of drinking. Figure 5A shows the connections that might develop in the configural network as a consequence of pairing consumption of the flavor with an injection of LiCl. The solid lines indicate unconditioned links, which may be excitatory or inhibitory, and the dashed lines indicate the connections formed during conditioning. In keeping with the principles embodied in Figure 2, the input unit for *F* will become rapidly connected to a configural unit, which will become more slowly connected to the output unit for illness. Subsequent presentations of *F* will then activate both output units, and because of the inhibitory link between them, drinking will be suppressed.

An advantage of allowing a US to activate an input unit is that it becomes possible, through further modification to the network, to explain the outcome of experiments in which the value of a US is modified after it has been used as a reinforcer for conditioning with a CS. The effect of this revaluation treatment is to alter the capacity of the CS to elicit a CR, and such an outcome is taken as evidence that conditioning promotes the growth of stimulus–stimulus, rather than stimulus–response, associations. In an experiment by Holland and Straub (1979), rats first received conditioning in which a noise CS was paired with food. The food was then paired with illness induced by an injection

of LiCl, and the effect of this devaluation treatment was to reduce considerably the strength of the CR elicited by the noise. Even if biological events can excite input units, this result would be difficult to explain, because it has always been assumed that configural units are connected to output units. Figure 5B shows the connections that would develop in this experiment, as the network has been developed thus far. The first two stages of the experiment will promote the growth of connections that will allow, first, noise to activate the food output unit and, second, food to activate the illness output unit. When the noise is presented for testing, the only output unit to be activated will be the one for food, and no effect of the devaluation treatment should be seen.

To explain the experimental findings, therefore, it is necessary to assume that the configural unit activated by noise will excite the input unit, rather than the output unit, for food (see Figure 5C). During the first stage of the experiment, noise will activate a configural unit, which will activate the input unit for food, which will then lead to a response by virtue of the unconditioned link between the input and the output units for food. Test trials with the noise, after reinforcer devaluation, will then activate both output units, and the CR it elicits will be weaker than it would have been if the devaluation treatment had been omitted. Of course, in Figure 5C, the configural unit activated by food might activate the input unit, rather than the output unit, for illness, but there is no evidence to permit a choice between these alternatives.

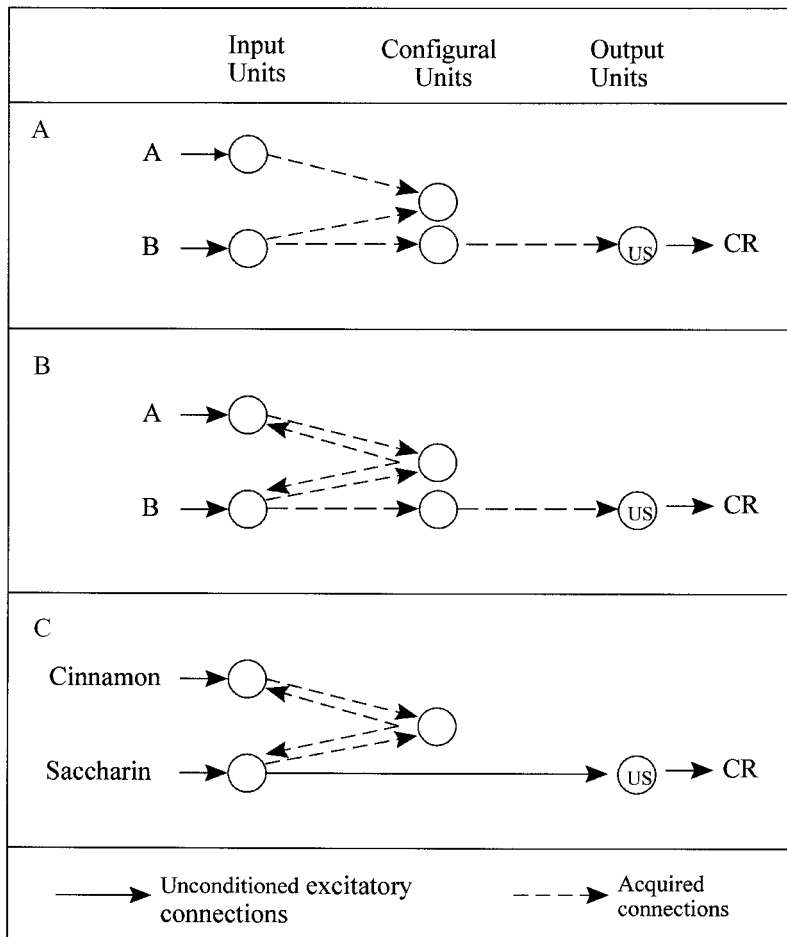
One consequence of the changes to the network that have been proposed concerns the effects of repeated pairings of a CS and a US. At first, the connections shown in Figure 5D will develop. Once this has happened, the CS will activate the input unit for food before food is presented, and the pattern of stimulation that now signals the US will be activation of the CS and US input units. Each of these units should then become connected to a configural unit (see Figure 5E). There are several important implications to the idea that a configural unit may partially be activated by the input unit for the US that it signals. For example, it is now possible for the network to explain the differential outcomes effect. Trapold (1970) provided rats with an instrumental discrimination in which responding on one lever resulted in a reward in the presence of a tone (T) and responding on a second lever resulted in a reward in the presence of a clicker (C). The discrimination was acquired more readily when the two responses led to qualitatively different rewards—food pellets and sucrose solution—than when they both led to the same reward. When responses lead to the same reward ( $R_1$ ), then as training progresses, the input units for C and  $R_1$  will acquire control over one response, and the input units for T and  $R_1$  will acquire control over the other response. Thus, the discrimination will effectively be between patterns  $R_1C$  and  $R_1T$ . But when the discrimination involves different rewards ( $R_1$  and  $R_2$ ) the discrimination will effectively be between  $R_1C$  and  $R_2T$ . Given that the trials for the former discrim-

ination involves the activation of a common input unit and those for the latter do not, the discrimination based on a single reward should be acquired more slowly.

A similar analysis can also be used to explain acquired equivalence and distinctiveness effects that have been found when two stimuli signal the same or different outcomes. Honey and Hall (1989) first gave one group of rats training in which two stimuli signaled food, A+ B+, while another group received an appetitive discrimination, A+ B-. Stimulus B was then paired with shock prior to test trials in which conditioned suppression in the presence of A was measured. The strength of this aversive response was stronger for the group that initially received the A+ B+ training. According to the foregoing analysis, A+ B+ training will result in A's and B's each exciting the input unit for food (F). Presenting B for aversive conditioning

will then enable the configural unit activated by BF to enter into an association with shock. When A is presented for testing, it will excite the input unit for food, which will partially activate the configural unit for BF and lead to an aversive CR. Such an effect would not be expected after training with A+ B-.

The proposals developed in this section also allow an explanation to be offered for such effects as mediated conditioning and extinction (Holland, 1981, 1990; Holland & Forbes, 1982). In a mediated conditioning study using rats (Holland, 1981), a CS signaled the delivery of distinctively flavored pellets. The CS was then paired on several occasions with an injection of LiCl. Subsequent tests revealed that this treatment reduced the attractiveness of the pellets originally associated with the CS. One explanation for this finding is that, as a result of the initial training, the



**Figure 6.** Possible modifications to the configural network proposed by Pearce (1994) to explain sensory preconditioning and flavor-flavor learning. (A) The connections that will develop during sensory preconditioning if nonreinforced exposure to a compound, AB, allows input units to be connected to a configural unit. (B) As for panel A, but with bidirectional links between the input and the configural units. (C) The connections that will develop as a result of drinking a mixture of cinnamon and saccharin.

presentation of the CS activated the input unit for the reinforcer with which it had been associated. The input units for the CS and for the distinctive food then became connected to a configural unit that entered into an association with food. When the CS was presented for the second stage, it again activated fully the configural unit, which now entered into an association with the effects of the injection of LiCl. Exposing the rats to the food in the test stage then partially activated the configural unit and excited an aversive CR that inhibited consumption.

An implication of this analysis is that the effect described by Holland (1981) will also be found when the CS is paired with footshock rather than with LiCl. This follows because the configural unit activated by input units for the CS and food should enter into an association with shock as readily as with illness. Although Holland (1981) failed to find any evidence of mediated conditioning when he replicated the above study by using shock, a more recent report by Ward-Robinson and Hall (1999) described successful mediated conditioning in these circumstances. One reason that it may be easier to demonstrate mediated conditioning between food and illness than between food and shock is that the response aroused by activating the input units for illness may be more likely to suppress eating than is the response aroused by activating the input units for shock.

### Sensory Preconditioning and Mediated Generalization

A potential problem for a configural theory of conditioning is posed by demonstrations of sensory preconditioning (Brogden, 1939). Rescorla and Durlach (1981) have described an autoshaping experiment in which pigeons received separate presentations of two simultaneous compounds, AB and CD, without reinforcement. Reinforced trials were then given with B and nonreinforced trials with D, which resulted, during later test trials, in stronger responding with A than with C.

As the configural network was originally formulated, the result of the experiment by Rescorla and Durlach (1981) might be that, first, the input units for A and B would be connected to a configural unit that would not acquire any associative strength. Next, after conditioning with B, the input unit for B would enter into an association with food via its own configural unit. Finally, when A was presented for testing, it would activate the AB configural unit, but this would not be able to activate either the input or the configural unit for B, and a CR would not be observed. The connections that would be predicted to be formed as a result of the training used by Rescorla and Durlach are shown in Figure 6A. For the sake of simplicity, the possibility considered earlier—that configural units might excite input rather than output units for the US—has been ignored.

To explain sensory preconditioning from the perspective of configural theory, Pearce (1987) proposed that after the training just described, a test trial with A would

activate a memory of AB, which would activate a memory of B, which would then activate a memory of food and would elicit a CR. However, no mechanisms were offered for this form of mediated generalization. One way of overcoming this omission, with reference to the configural network, is to allow the connections between the input layer and the configural layer to be bidirectional. Figure 6B shows the same network as the one in Figure 6A, except that the links between the input and the configural units have been modified in this way. If A should be presented for a test trial, it would activate the configural unit for AB, which would now activate the input unit for B through the bidirectional connection between them. Activation of the input unit for B would activate the output unit for the US and would generate a CR. The strength of this response would depend on the degree to which activation from A decreased as it worked its way through the network to the output unit. When A was presented, it would activate the configural unit for AB to a level given by the value of  ${}_A S_{AB}$ . This level of activation could then be multiplied by the similarity of B to AB,  ${}_B S_{AB}$ , and by a parameter,  $\gamma$ , in order to determine the degree of activation of the input unit for B. If the strength of the association between the configural unit for B and the US was  $\lambda$ , the level of activation of the output unit would be  ${}_A S_{AB} * {}_B S_{AB} * \gamma * \lambda$ . The parameter,  $\gamma$ , allows for the possibility that the influence of the connection from the configural unit to an input unit is not the same as that for the reciprocal link from the input unit to the configural unit.

An assumption that is fundamental concerning the network shown in Figure 2 is that connections between input and configural units will grow rapidly, which implies that not many exposures to a compound of two neutral stimuli will be required for successful sensory preconditioning. Support for this prediction can be found in Rescorla and Durlach (1981), who argue, on the basis of their experimental findings, that sensory preconditioning requires only a few exposures to the compound before it is effective. Indeed, they claim that the bulk of learning about the compound occurs on the first exposure to it. A further assumption has been that configural units will be formed for stimuli that are presented simultaneously, but not for those presented successively. It would thus follow that sensory preconditioning will be more effective with compounds of stimuli presented simultaneously than with those presented in succession. Support for this prediction can again be found in experiments cited by Rescorla and Durlach (1981; see also Rescorla, 1981).

Of course, sensory preconditioning is also possible with stimuli that are presented sequentially during their preexposure (Brogden, 1939; Pfautz, Donegan, & Wagner, 1978; Rizley & Rescorla, 1972). There are two ways configural theory might explain this result. If A is followed by B for the preexposure trials, A might excite a configural unit, which could then enter into an association with B. Associations involving a configural unit with another event are assumed to develop gradually over trials. It therefore fol-



lows that presenting stimuli sequentially will require more preexposure trials for successful sensory preconditioning than will presenting them simultaneously. Alternatively, if two stimuli are presented sequentially—A followed by B—the termination of A might leave a gradually decaying level of activity in its input unit. This activity could be sufficient to allow the input units for A and B to become connected to the same configural unit, with the link from A being weaker than that from B. The existence of the weak link between the unit input for A and the configural unit will then restrict severely any mediated generalization between A and B. Both of these explanations for sensory preconditioning imply that it will be a relatively weak phenomenon when the stimuli are presented sequentially, which, as was noted above, is consistent with experimental findings (see Rescorla & Durlach, 1981).

**Flavor-Flavor Learning and Bidirectional Links Between Input and Configural Units**

Further support for the idea that the connections between input units and configural units are bidirectional comes from experiments in which animals have been allowed to drink a compound consisting of a neutral flavor (say, cinnamon) and an attractive flavor without nutritional value (say, saccharin). The preference for cinnamon can be shown to increase as a result of this treatment (Capaldi, 1992; Holman, 1975). Figure 6C shows that presenting cinnamon and saccharin in the same solution will enable their input units to become connected to a configural unit. Because of its innate attractiveness, saccharin will already be con-

nected to an output unit, activation of which will be responsible for sustaining consumption. Because of the bidirectional links, subsequent presentation of cinnamon will activate the configural unit, which will activate the input unit for saccharin, which will activate the output unit and lead to increased consumption of cinnamon.<sup>4</sup> A number of predictions follow from this explanation that are in keeping with those derived for sensory preconditioning. One prediction is that it will require only one or two exposures to the compound for effective flavor-flavor learning. Evidence in support of this prediction comes from experiments by Boakes, Rossi-Arnaud, and Garcia-Hoz (1987) and Mehiel and Bolles (1988). Another prediction is that flavor-flavor learning is more likely to be effective with simultaneous presentations than with successive presentations of the flavors. Evidence supporting this prediction can be found in experiments by Holman (1975) and Lavin (1976). Finally, an assumption on which the network shown in Figure 2 is based is that once an input unit is connected to a configural unit, the strength of that connection will be unaffected by subsequent experience. For instance, if two stimuli, A and B, should become connected to a configural unit, repeated presentation of A by itself will have no impact on its connection with the configural unit, or on the connection between B and the configural unit. In other words, associations between simultaneously presented flavors will be extremely resistant to extinction. Evidence supporting this prediction comes from Capaldi, Myers, Campbell, and Sheffer (1983). They found that drinking a solution of wintergreen mixed with saccharin

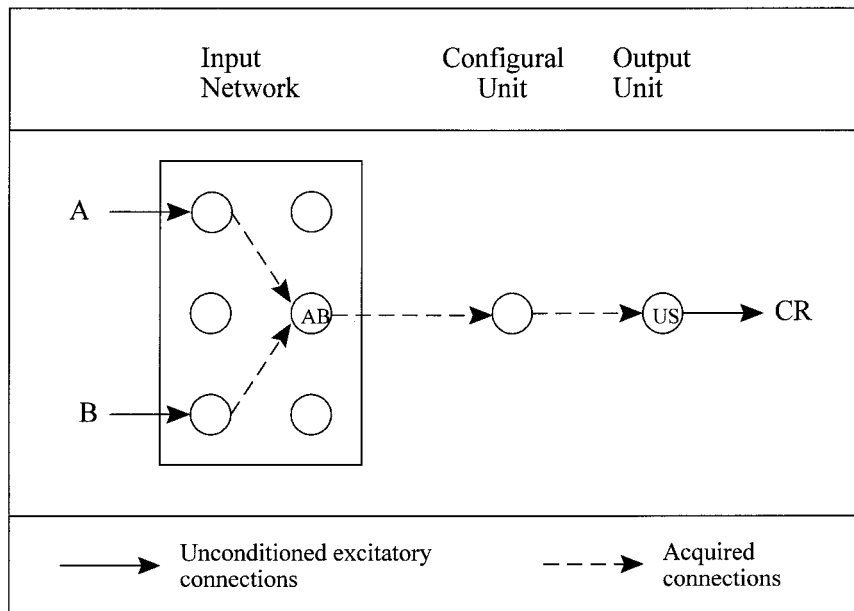


Figure 7. An extended input network to take account of the proposal of Rescorla and Durlach (1981) that perceptual differentiation may be necessary before a compound is treated as being distinct from its components.

resulted in a sustained increase in the preference for wintergreen over 28 days of testing.

A potential problem for the foregoing explanation is that sensory preconditioning can be disrupted by exposure to individual components of the compound (Rescorla, 1981; Rescorla & Durlach, 1981). Rescorla and Durlach have described an experiment by Freberg (1979), who found that sensory preconditioning consequent upon drinking a solution of salt and quinine was attenuated by exposure to quinine either before or after exposure to the compound (see also Westbrook et al., 1995). If sensory preconditioning depends on the formation of connections between the input units for salt and quinine and a configural unit, then, as was noted above, separate exposure to quinine should have no impact on these connections, and, thus, no impact on sensory preconditioning. It is worth noting that sensory preconditioning is not necessarily adversely affected by separate presentations of components of the compound. Rescorla and Durlach have described an experiment in which conditioning was conducted with one stimulus, A, before the compound, AB, was paired with the same US. Subsequent testing revealed that responding to B alone was influenced by the current associative properties of A. Thus, despite the conditioning trials with A by itself, sensory preconditioning remained effective (see also Speers, Gillan, & Rescorla, 1980).

Rescorla and Durlach (1981) explained the findings of Freberg (1979) by suggesting that when a compound, AB, is first presented, it is perceived as a unitary event. Subsequent conditioning with A will then activate the representation of AB and will enable it to enter into an association with the reinforcer. If B should then be presented, it will also activate the representation of AB and will elicit a CR. However, if animals should receive extensive exposure to B alone, a process of perceptual differentiation will take place that will lead to its activating a representation of itself, rather than of AB. Subsequent conditioning with A may then allow the representation of AB to enter into an excitatory association, but the reduced capacity of B to activate the representation of AB means that it will fail to excite a substantial CR.

There may, therefore, be two mechanisms for sensory preconditioning. One could be based on the principles shown in Figure 6B and might be particularly effective with compounds formed from distinctive components. Sensory preconditioning based on this mechanism would

not be seriously affected by exposure to the components of the compound. The other mechanism could be based on the principles proposed by Rescorla and Durlach (1981) and would apply principally to compounds whose elements initially are hard to distinguish. In terms of configural theory, the proposals of Rescorla and Durlach imply that the input layer must be expanded to an input network (see Figure 7). When a compound, such as AB, is presented to the network, its components will activate, either individually or together, a single output unit of the input network. If activation of this unit is followed by a US, it will enter into an excitatory association through a configural unit. However, repeated exposure to one element by itself will result in its exciting its own output unit to the input network, rather than the output unit for AB. Once this has occurred, responding in the presence of A will be unaffected by any training that takes place with B. Rather little is known about the factors that promote a perceptual differentiation of a component from a compound. In addition to exposure to the elements by themselves, another possibility is that it is more likely to take place when a compound and its components signal different outcomes, rather than the same outcome (see Pearce & Redhead, 1999).

### The Discrimination of Structure

A serious problem that confronts the configural network shown in Figure 2, as well as the elemental network shown in Figure 1, is posed by the ability of animals to solve what, at first sight, seems a relatively simple discrimination. Figure 8 shows two patterns that might be used in a discrimination in which food is signaled by the combination of horizontal light stripes to the left of vertical dark stripes (left-hand rectangle), but not by vertical light stripes to the left of horizontal dark stripes (right-hand rectangle). Both patterns are made up of the same set of features, but it is the way in which the features are put together, or structured, that determines the solution to the discrimination. For this reason, it will be referred to as a structural discrimination. If the left-hand pattern is presented to a configural network based on the principles of the network in Figure 2, it will be decomposed into its constituent features—light, dark, horizontal, and vertical—which will become connected to a configural unit that is associated with food. Should the right-hand pattern be presented to the network, it will be decomposed into

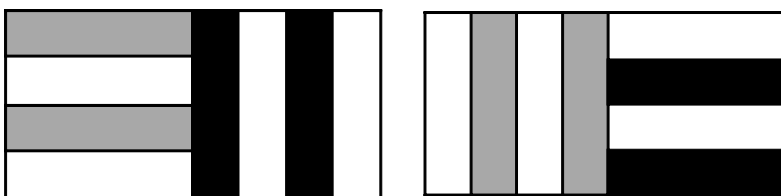


Figure 8. Two patterns that form the basis of a structural discrimination.

the same features, it will excite the same configural unit, and the discrimination will be insoluble. In fact, George and Pearce (2002) tested a group of pigeons with a structural discrimination based on the patterns shown in Figure 8, but using red and green instead of light and dark stripes, and the birds were able to solve this problem.<sup>5</sup> To solve the task, the pigeons had to appreciate that it was the combination of, say, red *and* horizontal together with green *and* vertical that signaled food.

One way in which a network based on the one shown in Figure 2 could solve a structural discrimination is to assume that the patterns are decomposed into coarser features than have been proposed. If the left-hand pattern in Figure 8 is broken down into two features of (1) light–horizontal-lines and (2) dark–vertical-lines, whereas its neighbor is broken down into (3) light–vertical-lines and (4) dark–horizontal-lines, the network would be able to solve the discrimination, because it involves four different features. Unfortunately, several adverse consequences follow from this proposal. One problem is that it leads to the prediction that generalization gradients will be unreasonably sharp. For instance, if conditioning is conducted with a red vertical line and then a test trial is conducted with a red horizontal line, these will be treated as entirely different features, and there will be no generalization from one to the other. Moreover, as Hummel and Biederman (1992) have noted, this type of coarse coding will lead to an unrealistically high number of input units that must code for every possible combination of color and orientation. In view of these problems, George, Ward-Robinson, and Pearce (2001) proposed a rather different modification to the configural network shown in Figure 2. In essence, they suggested that patterns are still broken down into their basic features, but that these features will excite two sorts of input units to the network. They will excite *simple* input units that are activated by the fundamental features of a pattern, and they will excite *structural* input units that are activated by pairs of features that occupy the same location. Thus, the pattern shown in the left-hand panel of Figure 8 will activate the simple input units of light, dark, horizontal, and vertical, together with two structural units that are sensitive to light–vertical-lines, and to dark–horizontal-lines, respectively. The complete set of input units will differ from the set that will be activated by the pattern on the right-hand side of Figure 8, and the configural network will be able to solve the discrimination. Just how a structural unit is formed so that it is sensitive to the combination of two inputs at the same location remains to be specified, but see Roelfsema, Engel, Konig, and Singer (1996) for one possible solution to this problem.

This modification to configural theory will not be sufficient for it to solve discriminations based on more sophisticated structural information than the co-occurrence of two features in the same location. For instance, Herrnstein, Vaughan, Mumford, and Kosslyn (1989) trained pigeons to discriminate between patterns in which a dot was either *inside* or *outside* a closed loop. This discrimination

was mastered with difficulty, and the pigeons failed to discriminate between novel exemplars that were dissimilar to the training exemplars. Thus, it is conceivable that the problem was solved not on the basis of the structural information of inside and outside, but by remembering each training pattern and its associated outcome. However, if the birds were genuinely relying on the structural relationships of inside and outside, the modification that has just been proposed would be unable to explain their success. Instead, a more complex network would be required. It is beyond the scope of the present article to describe such a network, which may need to be considerably more sophisticated than the network that has been presented thus far (Hummel & Biederman, 1992).

### CONCLUDING COMMENTS

A major concern of the present article has been to choose between a configural and an elemental account of the associations that are formed during a conditioning trial. According to elemental theories of conditioning, the pattern of stimulation provided by the signal for the US will activate a set of elements, each of which has the potential for entering into an association with the US. Configural theories are based on the supposition that some representation of this pattern of stimulation in its entirety will become connected with the US. These different assumptions have been shown to lead to different predictions about the outcomes of a variety of experimental tasks. Many of the tests of these predictions have led to findings that are consistent with a configural analysis of conditioning, although certain demonstrations of summation remain to be explained convincingly by this type of theory. The tests have also led to findings that challenge an elemental analysis based on the influential Rescorla–Wagner (1972) theory. However, a distinction between elemental and configural theories becomes harder to draw when certain elaborations of the Rescorla–Wagner theory are considered. For instance, Rescorla and Coldwell (1995) proposed that when two or more stimuli are presented together, they may interact perceptually so that a stimulus presented in one context would function as if it were a different stimulus when it is presented in a different context. One benefit of this proposal is that such effects as summation need not be an automatic consequence of presenting two stimuli together after they have been conditioned separately. If there is a perceptual interaction between them, summation might not be seen. An alternative and ingenious elaboration to the Rescorla–Wagner theory is the replaced-elements theory of Wagner and Brandon (2001). They proposed that two stimuli might interact by inhibiting some elements of the array that each of them excites when presented in isolation. Moreover, each inhibited element is assumed to be replaced by an element that is unique to the compound. The advantage of this theory is that it is able to make predictions that are consistent with either the Rescorla–Wagner theory or configural theory, depending

on the proportion of elements that are inhibited when two stimuli are presented together.

A problem that confronts both of these elaborations of the Rescorla–Wagner (1972) theory is that there is some uncertainty about the conditions that determine whether or not a perceptual interaction will occur between two stimuli or whether one stimulus will inhibit a large or a small proportion of the elements of another stimulus. Until these conditions can be specified, both elaborations will be in the unsatisfactory position of being able to explain almost any experimental outcome. From the evidence reviewed in this article, it does not appear likely that it will be possible to develop a simple, straightforward account of these conditions.

Not all of the findings considered in this article can be explained by the replaced-elements theory of Wagner and Brandon (2001), but this does not necessarily mean that it is impossible to explain them by an elemental theory of conditioning. Wagner and Brandon also put forward an inhibited-elements theory, the predictions of which are largely the same as those made by configural theory. It may ultimately, therefore, be difficult to discriminate between the different theoretical assumptions on which elemental and configural theories are based by referring to experimental evidence. Even if this should prove to be the case, there still remains a need to explore the issues considered in this article. It is clear that the associative properties of a CS can suffer a generalization decrement when the stimuli that accompany it are changed. At present, the factors that determine the degree of this decrement are by no means fully understood. Until they are understood, it will be difficult to make predictions with any theory about the effects of testing a stimulus in a different context to that used for training. There is also a need to understand fully the role of similarity in discrimination learning.

Configural theory is a relatively simple theory. It incorporates a single learning rate parameter,  $\beta$ , whose value has little impact on the asymptotic predictions made by the theory for a wide range of experimental tasks.<sup>6</sup> It also contains a parameter that specifies the similarity of one compound to another. A formal rule has been provided for calculating the value of this parameter. The theory thus makes unambiguous predictions, and it has the merit of being open to refutation. The simplicity of the theory, together with the clear predictions that can be derived from it, suggests that it may well be of value for guiding future research into the mechanisms of both generalization and discrimination in associative learning.

## REFERENCES

- ATKINSON, R. C., & ESTES, W. K. (1963). Stimulus sampling theory. In R. D. Luce, R. B. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3, pp. 121-268). New York: Wiley.
- AYDIN, A., & PEARCE, J. M. (1994). Prototype effects in categorization by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, **20**, 264-277.
- AYDIN, A., & PEARCE, J. M. (1995). Summation in autoshaping with short- and long-duration stimuli. *Quarterly Journal of Experimental Psychology*, **48B**, 215-234.
- AYDIN, A., & PEARCE, J. M. (1997). Some determinants of response summation. *Animal Learning & Behavior*, **25**, 108-121.
- BELLINGHAM, W. P., GILLETTE-BELLINGHAM, K., & KEHOE, E. J. (1985). Summation and configuration in patterning schedules with the rat and rabbit. *Animal Learning & Behavior*, **13**, 152-164.
- BOAKES, R. A., ROSSI-ARNAUD, C., & GARCIA-HOZ, V. (1987). Early experience and reinforcer quality in delayed flavor-food learning in the rat. *Appetite*, **9**, 191-206.
- BRANDON, S. E., & WAGNER, A. R. (1998). Occasion setting: Influences of conditional emotional responses and configural cues. In N. A. Schmajuk & P. C. Holland (Eds.), *Occasion setting: Associative learning and cognition in animals* (pp. 343-382). Washington, DC: American Psychological Association Press.
- BROGDEN, W. J. (1939). Sensory preconditioning. *Journal of Experimental Psychology*, **25**, 323-332.
- BURKE, C. T., & ESTES, W. K. (1957). A component model for stimulus variables in discrimination learning. *Psychometrika*, **22**, 133-145.
- CAPALDI, E. D. (1992). Conditioned food preferences. *Psychology of Learning & Motivation*, **28**, 1-33.
- CAPALDI, E. D., MYERS, D. E., CAMPBELL, D. H., & SHEFFER, J. D. (1983). Conditioned flavor preferences based on hunger level during original flavor exposure. *Animal Learning & Behavior*, **11**, 107-115.
- DARBY, R. J., & PEARCE, J. M. (1995). Effects of context on responding during a compound stimulus. *Journal of Experimental Psychology: Animal Behavior Processes*, **21**, 143-154.
- 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.
- DELAMATER, A. R., SOSA, W., & KATZ, M. (1999). Elemental and configural processes in patterning discrimination learning. *Quarterly Journal of Experimental Psychology*, **52B**, 97-124.
- ESTES, W. K., & HOPKINS, B. L. (1961). Acquisition and transfer in pattern-component discrimination learning. *Journal of Experimental Psychology*, **61**, 322-328.
- FREBERG, L. (1979). *Evidence supporting a perceptual learning model of sensory preconditioning of flavors*. Unpublished doctoral dissertation, University of California, Los Angeles.
- FRIEDMAN, M. P., & GELFAND, H. (1964). Transfer effects in discrimination learning. *Journal of Mathematical Psychology*, **1**, 204-214.
- GARCIA, J., & KOELLING, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, **4**, 123-124.
- GEORGE, D. N., & PEARCE, J. M. (2002). *Discrimination of structure: II. Implications for theories of object recognition*. Manuscript submitted for publication.
- GEORGE, D. N., WARD-ROBINSON, J., & PEARCE, J. M. (2001). Discrimination of structure: I. Implications for connectionist theories of discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes*, **27**, 206-219.
- GLUCK, M. A., & BOWER, G. H. (1988). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, **117**, 225-244.
- GULLIKSEN, H., & WÖLFLE, D. L. (1938a). A theory of learning and transfer: I. *Psychometrika*, **3**, 127-149.
- GULLIKSEN, H., & WÖLFLE, D. L. (1938b). A theory of learning and transfer: II. *Psychometrika*, **3**, 225-251.
- HAMM, R. J., PORTER, J. H., & OSTER, G. D. (1978). Compound stimulus control of operant, but not adjunctive, behavior. *Bulletin of the Psychonomic Society*, **12**, 167-170.
- HENDRY, J. S. (1982). Summation of undetected excitation following extinction of the CER. *Animal Learning & Behavior*, **10**, 476-482.
- HERRNSFEIN, R. J., VAUGHAN, W., JR., MUMFORD, D. B., & KOSSLYN, S. M. (1989). Teaching pigeons an abstract relational rule: Insideness. *Perception & Psychophysics*, **46**, 56-64.
- HINTZMAN, D. L. (1986). "Schema abstraction" in a multiple trace memory. *Psychological Review*, **93**, 411-428.
- HOLLAND, P. C. (1981). Acquisition of representation-mediated conditioned food aversions. *Learning & Motivation*, **12**, 1-18.
- HOLLAND, P. C. (1990). Event representation in Pavlovian conditioning: Image and action. *Cognition*, **37**, 105-131.
- HOLLAND, P. C., & FORBES, D. T. (1982). Representation-mediated ex-

- tion of conditioned flavor aversions. *Learning & Motivation*, **13**, 454-471.
- HOLLAND, P. C., & STRAUB, J. J. (1979). Differential effects of two ways of devaluing the unconditioned stimulus after Pavlovian appetitive conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, **5**, 65-78.
- HOLMAN, E. W. (1975). Immediate and delayed reinforcers for flavor preferences in rats. *Learning & Motivation*, **6**, 91-100.
- HONEY, R. C., & HALL, G. (1989). Acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Processes*, **15**, 338-346.
- HULL, C. L. (1943). *Principles of behavior*. New York: Appleton-Century-Crofts.
- HULL, C. L. (1945). The discrimination of stimulus configurations and the hypothesis of neural afferent interaction. *Psychological Review*, **52**, 133-139.
- HULL, C. L. (1949). Stimulus intensity dynamism (V) and stimulus generalization. *Psychological Review*, **56**, 67-76.
- HULL, C. L. (1952). *A behavior system*. New Haven, CT: Yale University Press.
- HUMMEL, J. E., & BIEDERMAN, A. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review*, **99**, 480-517.
- JAMES, J. H., & WAGNER, A. R. (1980). One trial overshadowing: Evidence of distributive processing. *Journal of Experimental Psychology: Animal Behavior Processes*, **6**, 188-205.
- KAMIN, L. J. (1969). Selective association and conditioning. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning* (pp. 42-64). Halifax: Dalhousie University Press.
- KEHOE, E. J. (1982). Overshadowing and summation in compound stimulus conditioning of the rabbit's nictitating membrane response. *Journal of Experimental Psychology: Animal Behavior Processes*, **8**, 313-328.
- KEHOE, E. J. (1986). Summation and configuration in conditioning of the rabbit's nictitating membrane response to compound stimuli. *Journal of Experimental Psychology: Animal Behavior Processes*, **12**, 186-195.
- KEHOE, E. J. (1988). A layered network model of associative learning: Learning to learn and configuration. *Psychological Review*, **95**, 411-433.
- KEHOE, E. J., HORNE, A. J., HORNE, P. S., & MACRAE, M. (1994). Summation and configuration between and within sensory modalities in classical conditioning of the rabbit. *Animal Learning & Behavior*, **22**, 19-26.
- KONORSKI, J. (1948). *Conditioned reflexes and neuron organization*. Cambridge: Cambridge University Press.
- KRUSCHKE, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.
- LAVIN, M. J. (1976). The establishment of flavor-flavor associations using a sensory preconditioning training procedure. *Learning & Motivation*, **7**, 173-183.
- LONG, C. K., & ALLEN, J. D. (1974). Stimulus compounding in pigeons. *Bulletin of the Psychonomic Society*, **4**, 95-97.
- MACKINTOSH, N. J. (1971). An analysis of overshadowing and blocking. *Quarterly Journal of Experimental Psychology*, **23**, 118-125.
- MACKINTOSH, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, **82**, 276-298.
- MAKI, W. S., & ABUNAWASS, A. M. (1991). A connectionist approach to conditional discriminations: Learning, short-term memory, and attention. In M. L. Commons, S. Grossberg, & J. E. R. Staddon (Eds.), *Neural network models of conditioning and action* (pp. 241-278). Hillsdale, NJ: Erlbaum.
- MCCLOSKEY, M., & COHEN, N. J. (1989). Catastrophic interference in connectionist networks: The sequential learning problem. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 24, pp. 109-165). San Diego: Academic Press.
- 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.
- MEHIEL, R., & BOLLES, R. C. (1988). Learned flavor preferences based on calories are independent of initial hedonic value. *Animal Learning & Behavior*, **16**, 383-387.
- MELTZER, D., & HAMM, R. J. (1976). Response summation in the pigeon. *Bulletin of the Psychonomic Society*, **7**, 515-518.
- MILLIER, P. J., & BEALE, I. L. (1977). Response summation to a compound stimulus in a context of choice. *Journal of the Experimental Analysis of Behavior*, **25**, 17-21.
- MYERS, K. M., VOGEL, E. H., SHIN, J., & WAGNER, A. R. (2001). A comparison of the Rescorla-Wagner and Pearce models in a negative patterning and a summation problem. *Animal Learning & Behavior*, **29**, 36-45.
- NAKAJIMA, S. (1997). Failure of inhibition by B over C after A+, AB-, ABC+ training. *Journal of Experimental Psychology: Animal Behavior Processes*, **23**, 482-490.
- NAKAJIMA, S., & URUSHIHARA, K. (1999). Inhibition and facilitation by B over C after A+, AB-, and ABC+ training with multimodality stimulus combinations. *Journal of Experimental Psychology: Animal Behavior Processes*, **25**, 68-81.
- PAGE, M. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral & Brain Sciences*, **23**, 443-512.
- PAVLOV, I. P. (1927). *Conditioned reflexes* (G. V. Anrep, Trans.). London: Oxford University Press.
- 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., ADAM, J., WILSON, P. N., & DARBY, R. J. (1992). Effects of discrimination training on responding during a compound conditioned stimulus. *Journal of Experimental Psychology: Animal Behavior Processes*, **18**, 379-386.
- PEARCE, J. M., AYDIN, A., & REDHEAD, E. S. (1997). Configural analysis of summation in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, **23**, 84-94.
- PEARCE, J. M., & GEORGE, D. N. (in press). Negative patterning with stimuli from three different dimensions. *Quarterly Journal of Experimental Psychology*.
- PEARCE, J. M., GEORGE, D. N., & AYDIN, A. (2002). Summation: Further assessment of a configural theory. *Quarterly Journal of Experimental Psychology*, **55B**, 61-73.
- PEARCE, J. M., GEORGE, D. N., REDHEAD, E. S., AYDIN, A., & WYNNE, C. (1999). The influence of background stimuli on summation in autoshaping. *Quarterly Journal of Experimental Psychology*, **52B**, 53-74.
- PEARCE, J. M., & HALL, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review*, **87**, 532-552.
- PEARCE, J. M., & REDHEAD, E. S. (1993). The influence of an irrelevant stimulus on two discriminations. *Journal of Experimental Psychology: Animal Behavior Processes*, **19**, 180-190.
- PEARCE, J. M., & REDHEAD, E. S. (1995). Supernormal conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, **21**, 155-165.
- PEARCE, J. M., & REDHEAD, E. S. (1999). Enhanced Pavlovian conditioning with a change in appetitive reinforcer. *Animal Learning & Behavior*, **27**, 369-378.
- PEARCE, J. M., REDHEAD, E. S., & GEORGE, D. N. (2002). Summation in autoshaping is affected by the similarity of the visual stimuli to the stimulation they replace. *Journal of Experimental Psychology: Animal Behavior Processes*, **28**, 175-189.
- PEARCE, J. M., & WILSON, P. N. (1990a). Configural associations in discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes*, **16**, 250-261.
- PEARCE, J. M., & WILSON, P. N. (1990b). Feature-positive discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes*, **16**, 315-325.
- PEARCE, J. M., & WILSON, P. N. (1991). Failure of excitatory condition-

- ing to extinguish the influence of a conditioned inhibitor. *Journal of Experimental Psychology: Animal Behavior Processes*, **17**, 519-529.
- PFAUTZ, P. L., DONEGAN, N. H., & WAGNER, A. R. (1978). Sensory preconditioning versus protection from habituation. *Journal of Experimental Psychology: Animal Behavior Processes*, **4**, 286-295.
- REDHEAD, E. S., & PEARCE, J. M. (1995a). Similarity and discrimination learning. *Quarterly Journal of Experimental Psychology*, **48B**, 46-66.
- REDHEAD, E. S., & PEARCE, J. M. (1995b). Stimulus salience and negative patterning. *Quarterly Journal of Experimental Psychology*, **48B**, 67-83.
- RESCORLA, R. A. (1969). Pavlovian conditioned inhibition. *Psychological Bulletin*, **72**, 77-94.
- RESCORLA, R. A. (1972). "Configural" conditioning in discrete-trial bar pressing. *Journal of Comparative & Physiological Psychology*, **79**, 307-317.
- RESCORLA, R. A. (1981). Simultaneous associations. In P. Harzem & M. D. Zeiler (Eds.), *Predictability, correlation, and contiguity* (pp. 47-80). Chichester, U.K.: Wiley.
- RESCORLA, R. A. (1989). Redundant treatments of neutral and excitatory stimuli in autoshaping. *Journal of Experimental Psychology: Animal Behavior Processes*, **15**, 212-223.
- RESCORLA, R. A. (1997). Summation: Assessment of a configural theory. *Animal Learning & Behavior*, **25**, 200-209.
- RESCORLA, R. A., & COLDWELL, S. E. (1995). Summation in autoshaping. *Animal Learning & Behavior*, **23**, 314-326.
- RESCORLA, R. A., & DURLACH, P. (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.
- RIZLEY, R. C., & RESCORLA, R. A. (1972). Associations in second-order conditioning and sensory preconditioning. *Journal of Comparative & Physiological Psychology*, **81**, 1-11.
- ROELFSEMA, P. R., ENGEL, A. K., KONIG, P., & SINGER, W. (1996). The role of neuronal synchronization in response selection: A biologically plausible theory of structured representations in the visual cortex. *Journal of Cognitive Neuroscience*, **8**, 603-625.
- RUMMELHART, D. E., HINTON, G. E., & WILLIAMS, R. J. (1988). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (Vol. 1, pp. 318-362). Cambridge, MA: MIT Press.
- SPEERS, M. A., GILLAN, D. J., & RESCORLA, R. A. (1980). Within-compound associations in a variety of conditioning procedures. *Learning & Motivation*, **11**, 135-149.
- SPENCE, K. W. (1936). The nature of discrimination learning in animals. *Psychological Review*, **43**, 427-449.
- SPENCE, K. W. (1937). The differential response in animals to stimuli varying within a single dimension. *Psychological Review*, **44**, 430-444.
- SPENCE, K. W. (1952). The nature of the response in discrimination learning. *Psychological Review*, **59**, 89-93.
- STRUB, H., & BARSKE, B. W. (1977). A stimulus-compounding assay of conditioned inhibition and excitation. *Learning & Motivation*, **8**, 414-428.
- SUTTON, R. S., & BARTO, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, **88**, 135-170.
- TRAPOLD, M. A. (1970). Are expectancies based on qualitatively different reinforcing events qualitatively different? *Learning & Motivation*, **1**, 129-140.
- WAGNER, A. R. (1969). Stimulus validity and stimulus selection in associative learning. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning* (pp. 90-122). Halifax: Dalhousie University Press.
- WAGNER, A. R. (1971). Elementary associations. In H. H. Kendler & J. T. Spence (Eds.), *Essays in neobehaviorism: A memorial volume to Kenneth W. Spence* (pp. 187-213). New York: Appleton-Century-Crofts.
- WAGNER, A. R. (1976). Priming in STM: An information-processing mechanism for self-generated or retrieval-generated depression in performance. In T. J. Tighe & R. N. Leaton (Eds.), *Habituation: Perspectives from child development, animal behavior, and neuropsychology* (pp. 95-128). Hillsdale, NJ: Erlbaum.
- 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. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 5-47). Hillsdale, NJ: Erlbaum.
- WAGNER, A. R., & BRANDON, S. E. (2001). A componential theory of Pavlovian conditioning. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 301-336). Mahwah, NJ: Erlbaum.
- 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). New York: Academic Press.
- WARD-ROBINSON, J., & HALL, G. (1999). The role of mediated conditioning in acquired equivalence. *Quarterly Journal of Experimental Psychology*, **52B**, 335-350.
- WEISS, S. J. (1977). The isolation of stimulus-reinforcer associations established with multiple schedules. *Animal Learning & Behavior*, **5**, 421-429.
- WESTBROOK, R. F., DUFFIELD, T. Q., GOOD, A. J., HALLIGAN, S., SETH, A. K., & SWINBOURNE, A. L. (1995). Extinction of within-event learning is contextually controlled and subject to renewal. *Quarterly Journal of Experimental Psychology*, **48B**, 357-375.
- WHITLOW, J. W., JR., & WAGNER, A. R. (1972). Negative patterning in classical conditioning: Summation of response tendencies to isolable and configural components. *Psychonomic Science*, **27**, 299-301.
- WILSON, P. N., & PEARCE, J. M. (1992). A configural analysis of feature-negative discrimination learning. *Journal of Experimental Psychology: Animal Behavior Processes*, **18**, 265-272.
- WILTZ, R. A., BOREN, J. J., MOERSCHBAECHER, J. M., CREED, T. L., & SCHROT, J. F. (1973). Generalization gradients and combined stimulus control after equal training with two related stimuli. *Psychological Reports*, **32**, 1003-1008.
- WOODBURY, C. B. (1943). The learning of stimulus patterns in dogs. *Journal of Comparative Psychology*, **35**, 29-40.
- YOUNG, D. B., & PEARCE, J. M. (1984). The influence of generalization decrement on the outcome of a feature positive discrimination. *Quarterly Journal of Experimental Psychology*, **36B**, 331-352.

## NOTES

1. Rescorla (1997) has reported a successful demonstration of summation, using a related experimental design. However, in his experiments, rats also received appetitive conditioning with a compound composed of a second auditory and a second visual stimulus. Pearce, George, and Aydin (2002) demonstrated that configural theory predicts that the conditioning trials with the compound will provide a source of excitation that will generalize to the test compound and will result in summation. In support of this prediction, Pearce et al. (2002) were able to replicate Rescorla's (1997) results when conditioning with the compound was included during the training stage, but not when it was omitted.

2. The additional trials with D, E, and F were included to make it difficult to use the number of stimuli on the screen as a cue for solving the discrimination.

3. I have conducted several computer simulations of the replaced-elements theory, based on the principles presented by Wagner and Brandon (2001), for the experimental design shown in the bottom row of Table 2. The outcome of each simulation was consistent with the results reported by Myers et al. (2001) but not with those reported by Redhead and Pearce (1995a).

4. If cinnamon should be presented in solution with a substance that has nutritional value, a more complex network will develop than the one shown in Figure 6C. In particular, the configural unit will gradually enter into an excitatory association with the nutritional consequences of consuming the solution.

5. In order to rule out theoretically uninteresting explanations for the solution of the discrimination, it was necessary to include the mirror images of the patterns shown in Figure 8 and to include additional patterns that incorporated a third color.

6. The design of the relative validity experiments conducted by Wagner, Logan, Haberlandt, and Price (1968) is one case in which the value of  $\beta$  is important for the outcome that is predicted by configural theory. In keeping with the Rescorla–Wagner (1972) theory, configural theory predicts the correct outcome to the experiments only if the value of  $\beta$  is greater for reinforced than for nonreinforced trials.

(Manuscript received October 1, 2001;  
accepted for publication February 28, 2002.)