

# Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning

GEOFFREY HALL

*University of York, York, England*

Pavlovian conditioning is taken to reflect the formation of links between the central representations of stimuli. A link will be formed when presentation of the relevant stimuli is scheduled in a way that ensures that two representations are activated concurrently. Once this has occurred, a representation can be activated not only by the occurrence of the appropriate stimulus but also by way of the link. Evidence is reviewed to suggest that activation produced by this second route is, in some ways, functionally equivalent to direct activation; in particular, an associatively activated representation (an *image*) appears capable of forming further associative links with other event representations. Learning about associatively activated stimulus representations may play a role in a range of phenomena. Its contribution to the following is discussed: sensory preconditioning, second-order conditioning, acquired equivalence and distinctiveness, equivalence class formation, and the perceptual learning effect. Finally, consideration is given to the way in which existing theories of associative learning might be modified in order to accommodate this process.

What may be termed the *standard model of classical conditioning* (Hall, 1994; Roitblat, 1987) holds that excitatory associations will be formed when the central representations of events are activated concurrently. These representations (or *nodes*; e.g., Wagner, 1981) are assumed, for the sake of theoretical convenience, to bear a direct relationship to the events described by the experimenter as stimuli. Thus, the node that corresponds to the conditioned stimulus (CS) will be activated when that event is presented to the animal; presentation of an unconditioned stimulus (US) will generate activity in a separate node corresponding to that stimulus. There is, however, a second route by which activity may be generated in a stimulus node. The existence of an associative link between nodes functions, according to our standard assumptions, to allow activation of one node produced, for example, by the presentation of the CS, to excite activity in the US node with which it is linked, in the absence of the occurrence of the US itself.<sup>1</sup> The presence of associatively generated activity in the US node is responsible for any conditioned response (CR) that the CS evokes. How this form of activity results in responding and what deter-

mines the exact form of the CR are interesting questions, but they will not be pursued here. Rather we shall focus on the question, What other properties, if any, does an associatively activated representation possess?

For the most part, theories of associative learning (e.g., Pearce & Hall, 1980; Rescorla & Wagner, 1972) have been able to function perfectly well without addressing this question at all. It has been enough to use the vigor of the CR as an index of the strength of the associative link, and there has been little need to say more about the precise nature of the activity that is produced by way of the link. In the absence of any specification to the contrary, therefore, we must assume that these theories accept the view that has been current in associative thinking for more than a century—that the activity induced by way of an associative link may be less intense than that produced by direct application of the US but will not differ from it in quality. Mill's (1843/1974, p. 852) statement of the laws of association described "secondary mental states" as simply being the original state in "inferior degree," "resembling the former but inferior in intensity." The implication is, therefore, that an associatively activated representation will have essentially the same properties as one that is directly activated by presentation of its own appropriate stimulus.

To assume that direct activation and associative activation produce the same sort of activity in a node may seem to make for simplicity in theorizing, but a problem immediately arises. As Konorski (1967) points out,

there is an essential *psychological* difference between these two events, because each gives rise to a distinct mental experience: whereas the activation of a set of gnostic units [a node] through the afferent pathway produces the experience of *perception* of a stimulus-object concerned, the ac-

---

Some of the experimental work reported here was supported by grants from the U.K. Medical Research Council and Biotechnology and Biological Sciences Research Council. The author thanks G. Alonso, C. Bonardi, I. Brugada, M. Symonds, and J. Ward-Robinson for helpful discussion. Correspondence should be addressed to G. Hall, Department of Psychology, University of York, York YO1 5DD, England (e-mail: gh1@york.ac.uk).

Note: This article is one of those occasionally invited by the editor in which authors have the opportunity to provide an overview of their research programs.—*Editor*

tivation of this set of units through associative pathways usually produces the experience of its *image*. (p. 170)

Konorski's solution to the problem is to point out that although direct activation may produce just the same state in a node as does activation via an associative pathway, direct activation also produces a range of other effects within the system that associative activation does not. Direct activation involves changes in the afferent system itself, along with activity elicited in the motor system concerned with orienting toward the source of the stimulation. It is this entire package of changes that constitutes perception; an image results when just one aspect (activity in the "gnostic unit") occurs in the absence of the others. In this way, Konorski is able to accommodate the distinction between perception and image while retaining the assumption that both involve the same sort of activity in the stimulus node. He is then able to go on and generate some interesting predictions about the conditions in which associations will be formed.

As we have noted, the standard assumption of associative theory is that excitatory connections will form between nodes that are activated concurrently. In most conditioning procedures, this will occur because two environmental events are presented simultaneously (or close together in time). But other patterns of activation are possible. In particular, presenting a stimulus that has been pretrained as a CS for some other event (S1) along with another, unrelated, stimulus (S2) will mean that direct activation of the node corresponding to S2 will coexist with associatively generated activity in node S1. Konorski's (1967) analysis, which makes no distinction between direct and associative activation, implies that this pattern of activity should result in the formation of excitatory links between the S1 and S2 nodes. If associative activation of the S1 node is reliably followed by S2, then an S1 → S2 association will be formed so that when S1 is actually presented, it will be able to evoke the "response proper to S2" (Konorski, 1967, p. 294). That is, the image of S1 should be able to function as a CS. In other circumstances (not fully specified by Konorski, but implicitly when the S2 precedes the associative activation of S1), the image will act as a US; that is, an excitatory association will be formed between S2 and the associatively activated S1 node. S2 will then be able to evoke the response proper to S1. In Konorski's terminology, S2 becomes a secondary CS, capable of evoking a secondary CR by virtue of its association with the S1 node that was activated originally by presentation of a primary CS (Konorski, 1967, p. 288).

An altogether different, and more radical, solution to the problem of distinguishing perception from image emerges from the theory of conditioning proposed by Wagner (1981; see also Mazur & Wagner, 1982; Wagner & Brandon, 1989). This distinguishes two qualitatively different states of activation that may be induced in a stimulus node. The primary state (A1) is activated only when the relevant stimulus is presented; if a node is activated associatively, the secondary (A2) state is generated. These

states differ in a number of ways. One important proposal is that type of response evoked by a node in the A2 state may well be quite different from that evoked by the A1 state. Another, with equally wide-ranging implications, is that the presence of A2 activity in a node will restrict the ability of the node to enter the A1 state when its stimulus is presented (it is as if the image interfered with perception). But for our present purposes, the most important difference between the two states is that they are assumed to differ in their ability to enter into new associative links. In contrast to Konorski (1967), Wagner's (1981) theory explicitly asserts that the A2 state (the image of an event) will not support the formation of excitatory associations. Rather, when one node is directly activated (in the A1 state) and another is associatively activated (in A2), an inhibitory link is formed, with the former acquiring the power to inhibit the latter. The formation and strengthening of excitatory links between a pair of nodes will occur only when A1 activation occurs concurrently in both.

We are faced, therefore, with clear alternative theories that propose quite different answers to the question (to use Konorski's terminology) of whether or not an image of event will serve as a substitute for direct perception of the event when it comes to association formation. The first two sections of this article review the experimental evidence that might help us to choose between the alternatives. They will show, to anticipate, that an image can indeed function as a CS in some circumstances and that it might also (although here the evidence is somewhat less compelling) be able to function as a US. In the next two sections, we consider the implication of these conclusions for some phenomena that have recently been the subject of experimental and theoretical attention: for acquired equivalence effects and for the processes responsible for perceptual learning. A final discussion concerns the wider theoretical issues raised by the proposal that conditioning can occur to an associatively activated stimulus representation.

### **IMAGE AS CS: SENSORY PRECONDITIONING AND RELATED PROCEDURES**

In the experiments to be described in this section, subjects experience presentations of a pretrained CS along with a new US, and evidence is sought for the formation of an excitatory link between the original associate of the CS (i.e., the associatively activated image of an event) and this US. This experimental procedure is essentially a version of the familiar sensory preconditioning design and we begin by considering this.

#### **Sensory Preconditioning**

The sensory preconditioning procedure involves two stages. In the first, two stimuli, A and B, are paired; they may be presented sequentially as A → B (e.g., Prewitt, 1967), or they may be presented as a simultaneous compound AB (e.g., Brogden, 1939; Rescorla, 1980). In the

second stage, B undergoes orthodox conditioning and is trained a CS signaling the occurrence of a motivationally significant US. A final test reveals the ability of A to evoke the CR conditioned to B in Stage 2 (see Table 1a). The standard interpretation of this result makes use of the notion of an associative chain. Stage 1 training, it is assumed, endows A with the ability to activate the representation of B; Stage 2 training endows B with the ability to activate the US representation. Thus, in the test stage, A is able to elicit the CR by way of the chain  $A \rightarrow B \rightarrow US$ .

Although the associative-chain account has no serious rival as an explanation for the sensory preconditioning effect produced by sequential (i.e.,  $A \rightarrow B$ ) Stage 1 training, there is reason to doubt its adequacy when Stage 1 consists of presentations of the simultaneous AB compound. The relevant finding comes from an experiment by Rescorla and Freberg (1978). They demonstrated that the sensory preconditioning effect was substantially reduced if, after training with the AB compound and before conditioning with B, the subjects were given experience of either A alone or B alone. That A-alone trials should have this effect is not surprising from one point of view—this procedure is likely to produce extinction of the  $A \rightarrow B$  link on which the associative-chain mechanism depends. It is not clear, however, why B-alone trials should have the same effect; the  $B \rightarrow A$  association may be weakened, but this is not taken to be an important link in the chain (but see Bolhuis & Honey, 1994). It seems that some process other than, or in addition to, the standard associative-chain mechanism plays a role in this form of sensory preconditioning.

One possibility (originally proposed in this context by Rescorla & Cunningham, 1978) is that, in the simultaneous sensory preconditioning procedure, the image of the test stimulus becomes a CS for the US. If we accept that exposure to the AB compound in Stage 1 will establish a  $B \rightarrow A$  association, then this means that the representation of A will be activated when B is presented on the conditioning trials of Stage 2. If an image can act as a CS, then the associatively activated representation of A will be able to form a direct association with the US at this stage. A will then be able to evoke the CR when the stimulus itself is actually presented in the test stage. The fact that the magnitude of the effect suffers when B-alone trials are interpolated between Stages 1 and 2 of training follows naturally from this analysis; these trials

will allow extinction of the  $B \rightarrow A$  association necessary for the image to be evoked during conditioning.

Simultaneous sensory preconditioning thus provides a first indication that conditioning can occur to an associatively activated representation. But this conclusion requires some qualification. First, it should not be supposed that this is the only, or even the main, source of the effect. The standard associative-chain mechanism could well play a role in simultaneous as well as sequential sensory preconditioning—certainly, the observation that A-alone presentations can undermine the sensory preconditioning produced by training with a simultaneous compound also seems to implicate the  $A \rightarrow B$  association in this case.<sup>2</sup> Second, although Rescorla and Freberg's (1978) results are grounds for doubting that the standard associative account is a complete explanation for simultaneous sensory preconditioning, they do not *require* us to accept that the effect involves the acquisition of associative strength by an image of the test stimulus. These authors (see also Rescorla, 1981; Rescorla & Durlach, 1981) propose an alternative. They suggest that exposure to the AB compound has the effect of establishing a unitary representation of this event that is perceptually similar to both A and B. Reinforcing B will thus endow the AB unit with associative strength that will generalize to A on test. Presentations of the stimulus elements alone (as in Rescorla & Freberg's, 1978, experiment) will disrupt the integrity of the unitary representation of AB formed in the first phase of training. As Rescorla (1981) acknowledges, many features of this account are inadequately specified. Nonetheless, the existence of this possibility weakens the case for concluding that conditioning with an image as the CS is, even in part, responsible for simultaneous sensory preconditioning. What is needed, given that simultaneous presentation of the stimuli is taken to be necessary for the formation of a unitary representation (Rescorla, 1981), is evidence from a procedure in which the critical stimuli are presented sequentially. The next section describes such a procedure.

### Backward Sensory Preconditioning

Table 1b presents a variant of the design for a standard experiment on sensory preconditioning with sequential presentations of the stimuli ( $A \rightarrow B$ ) in Stage 1. As usual, one of these stimuli is then trained as the CS for a motivationally significant US in Stage 2, and, in Stage 3, the ability of the other to evoke the CR is assessed. It differs from the standard procedure only in that it is the stimulus that was presented second in Stage 1 (B in this case) that is paired with the US in Stage 2 and the first (A) that is presented on test. For this reason, the procedure can be regarded as constituting a *backward* version of sensory preconditioning. (This term is intended to be purely descriptive; whether or not the procedure involves the formation of backward associations is an issue that will be taken up later.)

As we have seen, the associative account of the forward sensory preconditioning effect produced by  $A \rightarrow B$  training, attributes it to an associative chain, the elements

**Table 1**  
**Two-Stage Conditioning Procedures**

	Stage 1	Stage 2	Test
(a) Sensory preconditioning			
	A $\rightarrow$ B (or AB)	B $\rightarrow$ US	A
(b) Backward sensory preconditioning			
	A $\rightarrow$ B	A $\rightarrow$ US	B
(c) Second-order conditioning			
	B $\rightarrow$ US	A $\rightarrow$ B	A

Note—A and B represent neutral stimuli; US = unconditioned stimulus; AB indicates the two stimuli presented as a simultaneous compound.

of which are the A→B association formed in Stage 1 and the B→US association formed in Stage 2. The backward procedure, which trains A→US in Stage 2, will not establish such a chain, and no effect is to be expected on these grounds. Nor will a Stage 1 training procedure in which the stimuli are presented sequentially establish the unitary representation, and so there is no reason to predict an effect from this source. But the proposal that the image of stimulus can act as a CS leads to a different conclusion. According to this notion, the training given in Stage 2 involves not only a pairing of A with the US but also a pairing of the US with the image of B (the B node being associatively activated by way of the A→B link formed in Stage 1). The formation of a direct B→US link would then allow B to evoke the CR on test. A satisfactory demonstration of backward sensory preconditioning would, therefore, provide powerful evidence that an image can function as a CS.

Early studies of the backward sensory preconditioning procedure were not encouraging; with one exception, they failed to find any effect (e.g., Brown & King, 1969; Coppock, 1958; Tait, Marquis, Williams, Weinstein, & Suboski, 1969). (Note that the exception, from an experiment by Silver & Meyer, 1954, can be explained in terms of the formation of an associative chain. In this experiment, the A→B trials given in the first stage were very closely spaced, allowing the possibility of forward conditioning between the B that ended one trial and the A that began the next. The resulting B→A association would permit B to control the CR by way of the chain B→A→US.) It remains the case, however, that null results can never be decisive—the experiments cited above may have failed to detect the backward sensory preconditioning effect simply because of an inappropriate choice of procedure and parameters. Indeed, more recent work, making use of substantially different procedures, has produced altogether more interesting results.

**Image of food as CS, nausea as US.** Holland (1981) reported the first of a series of experiments designed to explore the possibility that an image (what he referred to as the representation of an event retrieved from memory) might serve as a substitute for the event itself in the formation of new associations (see also Holland, 1983, 1990; Holland & Forbes, 1982a, 1982b). Table 2 outlines the essential features of the training given to the various groups in Holland's (1981) Experiment 1. In Stage 1, rats in the critical experimental group (Group E in the table) received orthodox appetitive Pavlovian conditioning in which the delivery of distinctively flavored sucrose pellets was signaled by a tone. In Stage 2, the rats received two trials, each consisting of a single presentation of the tone followed, within the next 5 min, by an injection of LiCl. Although the detailed procedure is very different from anything discussed so far (in particular, the use of a motivationally significant event in the first stage of training is an innovation), it will be evident that this design constitutes an example of backward sensory preconditioning, with the tone as A and the food as B. In the test stage, the rats were offered 100 flavored pellets in

their home cages. As Table 2 shows, subjects in Group E consumed significantly fewer of the pellets than did subjects in the control groups that had not experienced the tone–food and food–LiCl pairings. Holland's (1981) interpretation of this result was as follows. Stage 1 training endows the tone with ability to evoke a representation of the food; this representation will be activated associatively in Stage 2 when it is followed by LiCl-induced nausea. The formation of an association between the representation and illness is shown by the tendency to reject this particular food in the test.

The effects generated in Holland's experiments were not large—the magnitude of the mediated-conditioning effect just described was, according to Holland (1990), less than 25% of that generated when equivalent training was given with the event itself rather than with its associatively activated representation. But this observation does not in itself challenge the suggestion that there may be functional identity between the states produced by direct activation and by associative activation; it may mean only that the image tends to be substantially less intense or salient than the percept. What is more, the principle of overshadowing supplies a reason why mediated conditioning is always likely to be a small effect. Holland's (1981) procedure required that the to-be-conditioned image (of food of a given flavor) be evoked by some external event (the tone trained in Stage 1). The occurrence of the LiCl-induced nausea is thus predicted by both of these events—the overt CS and the image of food—allowing the possibility that the presence of the former will restrict the acquisition of strength by the latter. The principle of selective association (e.g., Garcia, Brett, & Rusiniak, 1989), whereby an internal malaise appears to be more readily associated with a flavor than with a distal cue, may act to limit any overshadowing but does not necessarily eliminate it.

Holland's (1981) experimental procedure was designed to take advantage of the phenomenon of selective association. Indeed, other procedures that failed to take such advantage also failed to demonstrate a mediated-conditioning effect. In his Experiment 3, Holland (1981) conducted a study analogous to that described above, differing only in that the US used in Stage 2 was footshock rather than LiCl. This procedure was effective in establishing a CR to the cue presented during this stage, but it had no influence on the rats' willingness to consume the food that had been associated with that cue in Stage 1. The failure

**Table 2**  
**Outline Design and Results of Experiment by Holland (1981)**

Group	Stage 1	Stage 2	Test Results (pellets consumed)
E	T→f	T→LiCl	71
C1	T→f	T→0	94
C2	T→0	T→LiCl	91
C3	f→T	T→LiCl	86

Note—E is the Experimental group; the other three groups constitute a variety of control (C) conditions. T represents a tone; f, a wintergreen-flavored sucrose pellet; LiCl, an injection of lithium chloride. The test result is the number of food pellets consumed (out of 100 offered) over the course of 10 min.

to observe mediated conditioning with this procedure raises doubts about the generality of the effect. To demonstrate that certain phenomena are to be found only in those learning processes involved in the “gut-defense system” (Garcia et al., 1989) would be a finding of interest in itself; however, from our present perspective, which is a concern to establish whether conditioning with an image of a CS might be a principle of importance in learning generally, this outcome would be rather disappointing. Accordingly, we (the work was done in collaboration with J. Ward-Robinson) have recently undertaken a series of experiments intended to provide evidence for mediated conditioning in procedures not involving nausea-induced conditioned food aversion.

**Image of food as CS, shock as US.** Holland’s (1981) failure to establish an aversion to the image of food using a shock US need not be taken to imply that no food-US association was formed. An alternative is that the behavior established by this association (some form of conditioned fear response) might fail to show itself in the measure used on the test (the number of food pellets consumed; see Jackson & Delprato, 1974). Accordingly Ward-Robinson (1994) conducted a study in which the training procedure followed that used by Holland (1981), but which made use of a different test procedure. The design is shown in Table 3. In Stage 1, rats received intermixed trials in which presentations of a 10-sec auditory cue were followed by the delivery of a distinctively flavored sucrose pellet (either “grape” or “banana,” allegedly), and presentations of a 10-sec visual cue were followed by a pellet of the other flavor. The pellets were delivered to the same food tray, and, after 24 trials with each cue, the rats were reliably approaching the food tray in response to the occurrence of either cue. In Stage 2, food deliveries were discontinued and the subjects received three sessions each consisting of two trials with one cue (A in Table 3) followed by footshock and consisting of two nonreinforced presentations of the other cue (B). In preparation for the final test stage, two sessions of pretraining were given. On each, a single response lever was inserted into the apparatus (to the left of the food tray in one session, to the right in the other), and subjects were permitted to earn 20 pellets on a continuous reinforcement schedule. One of the levers yielded pellets of the type (x in Table 3) that had been associated with the reinforced stimulus of Stage 2; the other yielded type y pellets. The test consisted of a single 20-min session in which both levers were available, and responding on them produced a food pellet of the appropriate type according to a variable interval 30-sec schedule.

**Table 3**  
**Design of Experiment by Ward-Robinson (1994)**

Stage 1	Stage 2	Test
A→x and B→y	A→sh and B→0	Lever 1→x and Lever 2→y

Note—A and B are auditory and visual cues; x and y are distinctively flavored food pellets.

The results for the test session (group mean total numbers of responses to the lever that yielded pellet x and the lever that yielded pellet y) are shown in Figure 1. It is evident that animals responded somewhat less readily on the lever that yielded the pellet that had been previously paired with the reinforced CS of Stage 2 than on the other lever. These results are both disappointing and encouraging. They are disappointing in that the effect observed was small and evident only under ideal conditions. [The effect shown in Figure 1 interacted statistically with which of the levers delivered the devalued pellet,  $F(1,14) = 5.87, p < .05$ , and proved to be significant only in subjects for whom the right lever delivered this pellet,  $t(7) = 2.55, p < .05$ .] We have conducted several other experiments of this same general type with changes in the details of the training or testing procedures, but, in all, the results have been the same: Having the image of food with a shock US results in only a small measurable loss of effectiveness by that food type. On the other hand, we are encouraged (especially given the previous results of Holland, 1981) that any effect at all can be observed. There are many possible reasons why the effect might be small (perhaps our two food types are difficult for the rat to discriminate, the appropriate number of reinforced trials in Stage 2 is a matter for guesswork, etc.). But that there *is* an effect allows us to draw the important conclusion that conditioning with an image as CS is not a phenomenon restricted to the combination of flavor as CS and illness as US; there are grounds for thinking, therefore, that it involves a process of more general importance.

**Image of a neutral stimulus as CS, shock as US.** A further possible reason for the small size of the effects seen in Figure 1 is, of course, that it requires the animal to form an association between cue and consequence that, even with real events rather than images, might be difficult to establish (Garcia & Koelling, 1966). Holland’s (1981) solution to this problem was to use poison-induced nausea as the US when the image of food was the CS. An alternative is to make use of a standard exteroceptive event (or rather, its image) as the CS when shock is being used as the US. This is the strategy we employed in the experiment outlined in Table 4 (Ward-Robinson & Hall, in press).

In the first stage of the study, all subjects received training with two pairs of neutral events, the first immediately preceding the presentation of the second. Over six sessions, they received 12 A–B trials and 12 C–D trials, enough we hoped, for associations to be formed (not something we could monitor by observing changes in overt responding, given the nature of the events used). In Stage 2, there were two trials of A followed by footshock, allowing the possibility that the representation of B activated by the presentation of A might form an association with the shock. No shock was presented with Stimulus C. The consequences of this training were assessed over a series of three test sessions, each consisting of three presentations of B and three of D. For the test trials, stimulus duration was increased to 30 sec. Figure 2 shows the ability of these stimuli to suppress ongoing, food-rewarded

responding. Both stimuli evoked some suppression that diminished over trials, but the effect was substantially and significantly greater for Stimulus B. [An analysis of variance (ANOVA) on the data shown in Figure 2 yielded a significant stimulus  $\times$  session interaction,  $F(2,62) = 3.36$ ,  $p < .05$ .] This result is entirely consistent with the notion that the associatively activated representation of Stimulus B was able to form an association with the shock US during Stage 2 of training.

The effect demonstrated in Figure 2 is gratifyingly more substantial than that seen in Figure 1 and thus lends support to the suggestion that appropriate cue-to-consequence relations are necessary for a powerful effect to emerge. This conclusion may be premature, however. The procedure that generated the results of Figure 2 was arrived at only after a good deal of preliminary work in which a range of parameters (e.g., stimulus durations and modalities, the number of Stage 2 trials, US intensity, etc.) was manipulated. The need to juggle parameters in this way did not surprise us, given that we had no solid information to go on as to how best to arrange the concurrent activation of an image and a US representation, something on which their association is deemed to depend. The small effect seen in Figure 1 may be the result, therefore, not of an intrinsic reluctance of an image of food to form an association with a shock US, rather it may simply be that the particular training parameters used in that study were less than optimal. Pilot work of the sort that led to the results of Figure 2 could well reveal a set of parameters that would generate a powerful mediated-conditioning effect, even with an image of food as the CS.

**The role of backward conditioning.** It was argued above that the backward sensory preconditioning effect is not to be explained in terms of the operation of an associative chain—that the  $A \rightarrow B$  and  $A \rightarrow US$  associations likely to be formed with this procedure do not establish the required connection between B and the US. But this is to ignore the possibility that excitatory backward conditioning might occur. If the  $A \rightarrow B$  trials of Stage 1 re-

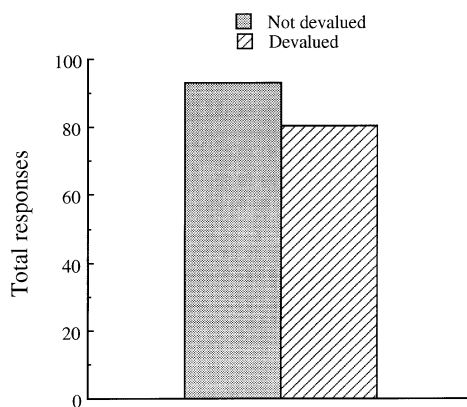


Figure 1. Group mean test scores for rats in the experiment by Ward-Robinson (1994). Two response levers (left and right) yielded different flavored food pellets, one of which had been devalued by prior training (see Table 3).

Table 4  
Design of Experiment by Ward-Robinson and Hall (in press)

Stage 1	Stage 2	Test
A $\rightarrow$ B and C $\rightarrow$ D	A $\rightarrow$ sh and C $\rightarrow$ 0	B and D

Note—A and C are auditory cues (a clicker and white noise) of 30-sec duration; B and D are visual cues (bright light or complete darkness) of 1-sec duration in Stage 1; sh represents shock.

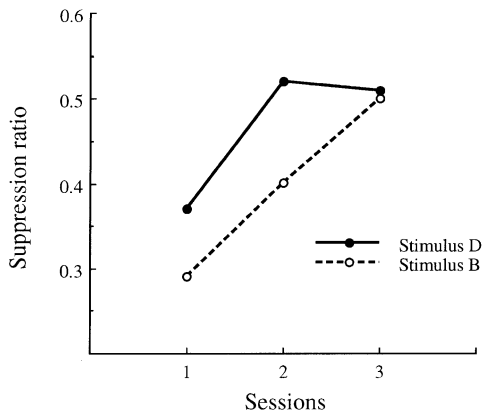
sulted in the formation of a  $B \rightarrow A$  association in which B acquired the ability to activate the representation of A, then the (forward) association of A with the US in Stage 2 would permit B to evoke the CR on test by way of the associative chain  $B \rightarrow A \rightarrow US$ . This analysis finds no place for conditioning with an image as CS. There are, however, reasons for doubting that this process is responsible for the backward sensory preconditioning effects just described.

First, although backward excitatory conditioning can undoubtedly occur (see Spetch, Wilkie, & Pinel, 1981, for a review), it does so only in a rather restricted set of conditions (e.g., it is usually necessary to give only a few training trials), and these conditions were not especially well met in the experiments described above. Second, Holland (1981) has directly addressed the role played by the associative chain (which, in his case, would be food  $\rightarrow$  tone  $\rightarrow$  US) in generating the effect that he obtains. If his result is a consequence of the formation of a food  $\rightarrow$  tone association during Stage 1 training with tone  $\rightarrow$  food, then it might reasonably be supposed that the effect would be especially likely if animals were given explicit food  $\rightarrow$  tone trials in this stage. This was just the training given to Control Group C3 (see Table 2) in Holland's (1981) Experiment 1, but these animals showed no sign of an aversion to the food after the tone had been paired with LiCl in Stage 2. The absence of an effect after training designed to allow the effective operation of the associative chain argues against the idea that such a chain is responsible for the effect seen in the parallel backward sensory preconditioning procedure, where the likelihood of such a chain being formed seems rather low (see also the study by Hall, Ray, & Bonardi, 1993, discussed below).

Given these arguments, the balance of the evidence suggests that the backward sensory preconditioning effect is probably not to be explained in terms of the operation of an associative chain. At any rate, it is clear that we need to take seriously the implications of the idea that the associatively activated representation of an event can acquire associative strength as a CS. Before exploring this matter, we first need to consider the evidence from a closely related set of experiments that bears on the issue of whether or not an image can function as a US.

#### IMAGE AS US: SECOND-ORDER CONDITIONING

The procedure for second-order conditioning is the same as that for sensory preconditioning, but with the



**Figure 2.** Group mean suppression ratios for test trials with Stimuli B and D from the experiment by Ward-Robinson and Hall (in press). The subjects had experienced B and D preceded by different signals in Stage 1. In Stage 2, the signal for B was followed by footshock (see Table 4).

order of the training phases reversed (see Table 1c). That is, subjects receive B→US in Stage 1 and A→B in Stage 2; the test phase shows that the stimulus that has not been directly paired with the US (in this case, A) is capable of evoking the CR. It is unsurprising, then, that the theoretical analysis of second-order conditioning should parallel quite closely what has been offered for sensory preconditioning.

First, second-order conditioning, like sensory preconditioning, is susceptible to explanation in terms of the formation of an associative chain. The procedure outlined above should be effective in establishing the associative chain A→B→US, and this supplies a ready explanation for the ability of A to evoke the CR when this stimulus is presented alone in the test phase. But again, as for sensory preconditioning, there are experimental results showing that this cannot be the whole story. The critical observation comes from the finding that, in many cases (e.g., Holland & Rescorla, 1975; Nairne & Rescorla, 1981; Rizley & Rescorla, 1972), the ability of A to elicit a second-order CR appears to be independent of the status of B. In particular, the second-order CR established in Stage 2 will often survive a phase of training in which B is presented alone—training that is effective in extinguishing the first-order CR and, thus, it may be assumed, in eliminating the B–US excitatory association that constitutes one of the links in the associative chain. It was this observation that led Konorski (1967) to develop his alternative account of second-order conditioning (and that forms the basis of his more general assertion that images can enter into associations). His proposal was that the second-order CR to A depends on the formation in Stage 2 of an excitatory association between A and the associatively activated representation of the US. But before we can accept second-order conditioning as evidence for conditioning with an image as the US, two problems must be dealt with. First, the procedures for second-order conditioning

can sometimes produce inhibitory learning; a satisfactory account of the phenomenon must be able to explain why this should be so. Second, we need to consider an alternative, S–R, explanation that has been offered for the results that cast doubt on the associative-chain hypothesis.

### Inhibitory Learning

The standard procedure for establishing a stimulus (A) as a Pavlovian conditioned inhibitor is to give training in which reinforced B trials are intermixed with nonreinforced trials with the AB compound. It is thus simply a variant of the second-order conditioning procedure. It differs in a number of details, but the only factor that has been established as determining whether the test stimulus will acquire inhibitory rather than excitatory properties is the amount of training given—excitation is observed early in training, but gives way to inhibition if training is prolonged (Yin, Barnett, & Miller, 1994). As Yin et al. observe, this result is consistent with the possibility that Stimulus B concurrently acquires both excitatory and inhibitory properties but that the development of the latter proceeds rather slowly so that only with extended training will it come to outweigh the effects of the former. To assume that excitatory and inhibitory learning are governed by different rate parameters is in no way contentious; the issue for the theorist is to specify just which features of the training situation generate excitation and which generate inhibition. In this, neither of the theories of major interest to us here (those of Wagner and of Konorski) is entirely successful.

The problem for Konorski (1967) is to explain how a second-order conditioning procedure could ever produce a CS that is a net inhibitor. The source of the excitation is, as we have seen, the formation of the excitatory association between the second-order CS and the US representation activated by the first-order CS. Inhibitory learning, according to Konorski, also depends on the formation of excitatory associations, but this time between the CS and a center activated by the absence of the US. This no-US center is assumed to have inbuilt inhibitory connection with the US center.<sup>3</sup> Although some details of the process are not fully specified, it appears to be assumed that activation of the no-US center occurs as a consequence of activation of the US center in the absence of the US itself. Thus, the second-order procedure should produce inhibitory as well as excitatory learning. What is less clear is why there should be a shift, over the course of training, in the relative strengths of the two associations so that the (second-order) excitation gives way to inhibition.

In this training procedure, it is supposed that strengthening of the CS–US and CS–no-US associations will go on together, and, without special assumptions, there is no obvious reason why the latter should overtake the former. Indeed, if we accept the plausible notion that the level of activity in the no-US representation is directly determined by that in the US representation (the level of disappointment or relief at the omission of the US being

set by the strength of the original expectation), then the most straightforward conclusion is that there will be no net change in the ability of the CS to evoke the CR. The basic theoretical difficulty becomes most starkly apparent when we consider simple acquisition and extinction. Once a CS-US association has begun to form in acquisition, the CS will be able to evoke activity in the US node that will further strength the connection. This will be true both during acquisition and in extinction when the US itself is omitted. Even if we allow that US activity might eventually give rise to activity in a no-US node, it is not clear that the inhibitory learning that would then occur would outweigh the excitation that is also developing—no asymptote would be reached, and conditioned responding would be present throughout extinction (see also Holland, 1983). Evidently, to accept Konorski's (1967) account of second-order conditioning will require a significant reworking of some other of the basic tenets of his theory as it applies to inhibitory learning.

On the other hand, Wagner's theory (e.g., Wagner & Larew, 1985) has no difficulty in predicting that the second-order training procedure should generate inhibitory learning—a central feature of his theory is the notion that a stimulus accompanying the associative activation of a US node in the absence of the US itself will acquire inhibitory properties. Explaining the excitatory component is more problematic. The suggestion that B acquires its power by way of the chain  $A \rightarrow B \rightarrow US$  can be only a partial solution, given the evidence (cited above) that, in many instances of second-order conditioning, no associative chain appears to be formed. For these, it is necessary to find some other source for the second-order CR. The possibility that this might be a process of S-R learning is discussed next.

### The S-R Alternative

As Rescorla (1973) has pointed out, activation of the US representation is not the only effect likely to be produced by the presentation of Stimulus B in Stage 2 of a second-order conditioning experiment. Since this stimulus has been established as a first-order CS, it will be capable of evoking a CR, and the occurrence of this response in the presence of A allows the possibility of a link being formed between S (Stimulus A) and R (the B-evoked CR). This account, like that proposed by Konorski (1967), has no difficulty in accommodating the fact that, once the second-order response has been established, the status of the first-order CS is immaterial.

It should be made clear at the outset that the relevant aspect of the CR in this form of learning need not be the overt response evoked by the first-order CS. Such an interpretation of S-R theory would require, for the relevant association to be formed, that the target response be elicited in the presence of Stimulus A. But there are experimental results that make it necessary to reject any such strict interpretation. In particular, Nairne and Rescorla (1981) have demonstrated perfectly good second-order conditioning with a first-order CS that evoked no obvious

CR. In several experiments, they demonstrated that pigeons will develop the response of pecking to a keylight when this stimulus is associated with a first-order CS (a diffuse cue) that evokes no such response. The S-R theory needs to assume that, in this case, the relevant response is not keypecking itself but some central emotional response or state evoked by the first-order CS as a consequence of its association with food. (It is possible that the second-order responding seen in some of these experiments is in part the result of the formation of an associative chain—see also Leyland, 1977—but this cannot explain the instances reported by Nairne & Rescorla in which the second-order CR survived extinction of the first-order CS.)

It is open to the theorist who accepts the framework proposed by Wagner (1981) to conclude that second-order conditioning depends on the formation of an association between the second-order CS and the emotional response produced when the US node is put into the A2 state by presentation of a first-order CS. But to do so is to take a significant step toward adopting an important element of the rival theory. Both this version of S-R theory and the version of S-S theory proposed by Konorski (1967) accept that the critical association is between the second-order CS and the associatively generated effects of the first-order CS; both also accept that these effects will include, importantly, the arousal of a central affective state or response. They differ only in that one supposes that the second-order CS comes to activate the US representation (arousal of the affective state being an intrinsic aspect of this activation), whereas the other supposes the association to be with a separate mechanism that is responsible for organizing the emotional response engendered by activation of the US representation.

Evidence that might allow us to discriminate between these alternatives comes from experiments (originally designed to distinguish between the S-R and associative-chain accounts) in which the value of the US is changed after the second-order conditioning procedure has been completed. Thus, for example, Holland and Rescorla (1975) established a second-order CR in rats with food as the primary US and then devalued the food by pairing it with high-speed rotation. Their finding—that the first-order CR was reduced in magnitude, whereas the second-order CR was not—was taken to show that the latter is not mediated by activation of the US representation. Results of this sort certainly suggest that a direct S-R association can play a role in some second-order conditioning procedures; however, it is another matter whether they show that Konorski's (1967) version of S-S theory should be rejected altogether. In the Holland and Rescorla (1975) experiment, the first-order CS evoked an overt CR, and the evidence for second-order conditioning was the ability of the second-order CS to evoke this same response. However, in other procedures in which there is no obvious overt CR to the first-order CS, S-R learning may be less important. We may postulate that the critical association in these preparations is between the second-order CS and a central emotional response, but, without evidence to the



contrary, the alternative—that the stimulus directly activates the US representation—remains a possibility.

### Conclusions

This review of some aspects of second-order conditioning leads to no very satisfactory conclusion. The procedure allows the possibility that a range of associations will be formed: between the second-order CS and the first-order CS, between the second-order CS and the CRs (overt or central) evoked by the first-order CS, and between the second-order CS and the associatively activated representation of the original US. We have good evidence that, in some circumstances, the S–R association will be formed and that, in others, the association between second-order and first-order CSs is not of importance. No doubt the particular association that dominates in any given training situation will be determined by how that situation influences the relative salience of the various possible associates (see Holland, 1985). What matters here, however, is the conclusion that there is no evidence that compels us either to accept or to reject the suggestion that one relevant association might be between the second-order CS and the image of the US. For the time being, therefore, we should continue to entertain the possibility that such associations can be formed and consider their implications for the phenomena to be discussed next.

### ACQUIRED EQUIVALENCE AND ACQUIRED DISTINCTIVENESS

It has long been held (e.g., since James, 1890) that associative learning processes might be capable of modifying the ease with which stimuli can be discriminated. The first clear statement of the idea in its modern form we owe to Miller and Dollard (1941; but see also Hull, 1939). In their analysis of what they called the *acquired equivalence of cues*, they suggested that a pair of cues would be rendered less discriminable by training in which each became associated with the same consequence. The complementary idea of *acquired distinctiveness* was made explicit a little later (Miller, 1948)—that training in which the cues become associated with different events will render them more readily discriminable. The aim of this section of the article is to evaluate the possibility that associative learning about images might constitute the mechanism of these effects. After a brief review of some basic demonstrations of the phenomena, an analysis in terms of image conditioning is offered. The application of this analysis to further, rather more complex, instances of acquired equivalence is then discussed.

#### Demonstration of the Phenomena

The form of training that Miller and Dollard (1941) had in mind in this context was one in which human subjects learned to apply verbal labels, the cues being given very different names in distinctiveness training but called by the same name in equivalence training. The 1950s and 1960s saw the publication of a substantial body of experimental work (reviewed by Hall, 1991) designed to dem-

onstrate that such training can modify the ability of the subjects to perform further tasks involving the same cues. With an appropriately chosen test task, the possibility of direct transfer from the first phase of training can be excluded (there is no reason, for instance, why having learned labels for a set of colored lights should directly influence the subjects' performance on a choice-reaction task involving these lights; see, e.g., Gagné & Baker, 1950). Accordingly, transfer effects can be taken to reflect changes in the discriminability of the cues. This research left a number of issues unresolved. Most critically, it failed to demonstrate that equivalence and distinctiveness training have their effects by virtue of the associations they establish (J. J. Gibson & E. J. Gibson, 1955, argued that the associations formed during discrimination training were irrelevant and that such training had its effects because it compelled the subject to attend to distinctive aspects of the stimuli). But the reality of the basic effect—transfer between discrimination tasks that appear to have nothing in common except the cues involved—was clearly established by this work.

A demonstration of the phenomenon in animal subjects came early, with the classic study by Lawrence (1949) of transfer from a simultaneous to a successive discrimination involving the same cues. His results were very influential (see, e.g., Sutherland & Mackintosh, 1971), but his interpretation that the transfer observed was a consequence of changes in the distinctiveness of the cues did not go uncontested (see, e.g., Siegel, 1967, who argued that specific response patterns acquired in the learning of a simultaneous discrimination could transfer to and influence performance on a successive discrimination). We will consider a more recent attempt to demonstrate the phenomenon in animals (Honey & Hall, 1989), which, by making use of classical conditioning procedures, may escape (at least some of) the problems that cause trouble in earlier work. At any rate, experiments using Honey and Hall's (1989) general procedure form the focus of the theoretical analysis to be presented later.

Table 5a presents the design of a three-stage experiment in which the subjects were rats. In Stage 1, all subjects received training with three auditory stimuli (a noise, A; and tone and clicker, counterbalanced, as Stimuli B and C). For all, B and C had different consequences; B was always followed by food, and C was not. The groups differed in the treatment given to Stimulus A. For Group A+/B+, the noise was reinforced; however, for Group A–/B+, it was not. For the first of these groups, then, B and A were followed by a common consequence, a training procedure supposed to induce equivalence; A and C, on the other hand, had different consequences in Stage 1, a procedure supposed to establish distinctiveness. For Group A–/B+, the reverse holds true, with A and C having the same treatment, and A and B having different treatments. The question of interest was whether this training would induce subjects to treat as equivalent the pair of stimuli that had had the same consequence and enable them to discriminate more readily between those that had been associated with different consequences.

**Table 5**  
**Acquired Equivalence and Distinctiveness:**  
**Experimental Designs**

	Stage 1	Stage 2	Test
(a) Honey and Hall (1989)			
Group A+/B+	B→f A→f C	A→sh	B and C
Group A-/B+	B→f A C	A→sh	B and C
(b) Honey (1990)	A→0 B→0	A→f	B and C
(c) Hall et al. (1993)			
Group +B/+A	f→B f→A C	A→sh	B and C
Group +B/-A	f→B A C	A→sh	B and C
(d) Bonardi et al. (1993)	X→A+ X→C+ Y→B+ Y→D+	A+ D-	B and C
(e) Bonardi and Hall (1994)	X(A+)/A- Y(B+)/B- C+/C-	A+	B and C

Note—Uppercase letters represent auditory or visual cues; f and + indicate food; sh indicates shock. X(A+) denotes reinforcement of A in the presence of X. In Bonardi and Hall (1994), the treatment given to C occurred in two separate phases, initial acquisition followed by extinction.

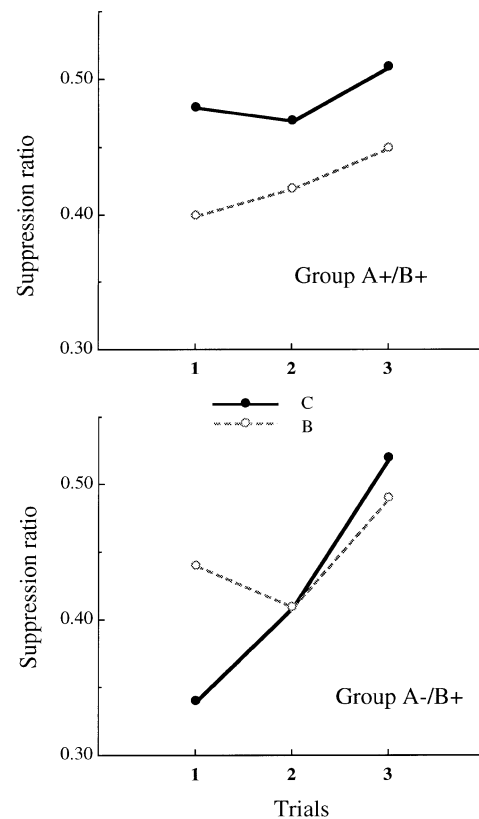
In order to reveal any such effects, the noise was given new conditioned properties in Stage 2 by being paired with shock. Stage 3 assessed the extent to which the effects of this training generalized to B and C. The results shown in Figure 3 come from a test in which the generalized tendency of B and C to evoke suppression of food-rewarded baseline response was assessed. Suppression was not profound, but, for both groups, there was significantly more suppression to the stimulus (B for Group A+/B+; C for Group A-/B+) that had received Stage 1 training equivalent to that given to the noise.

It is difficult to explain this pattern of results in terms of the transfer of specific responses acquired during Stage 1. Certainly, those stimuli given food-reinforced training in that stage will acquire the power to evoke CRs (e.g., approaching the site of food delivery), and these could be evoked by the (previously food-reinforced) Stimulus B in the test stage. One could postulate that this CR might tend to compete with the suppressive tendencies acquired by generalization from A, but such an effect would apply only to the results for Group A-/B+ in which B evoked less suppression than C. For Group A+/B+, by contrast, it was Stimulus B that evoked more suppression. Accordingly, the transfer evident in this experiment must come from a different source. Honey and Hall (1989) concluded that it reflected training-induced changes in the extent of generalization between the stimuli—the ac-

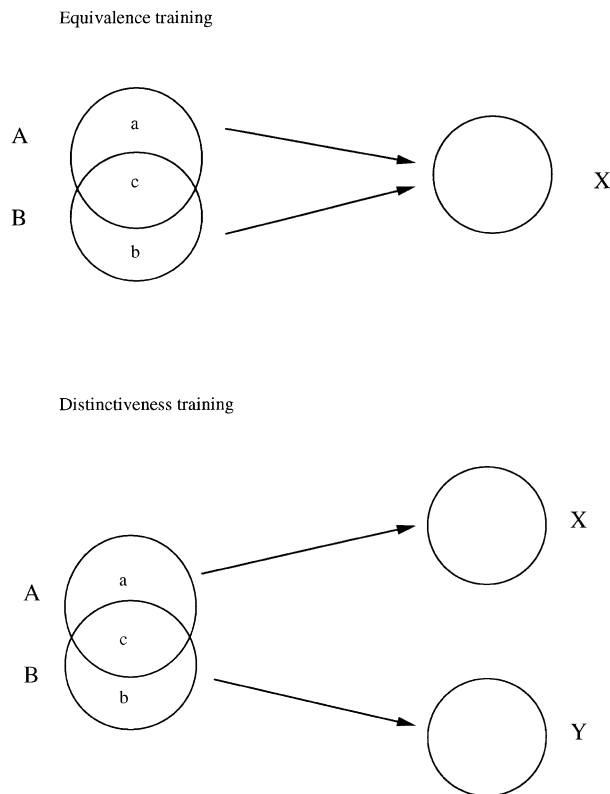
quisition of equivalence and/or distinctiveness. As with most experiments that have succeeded in demonstrating equivalence/distinctiveness effects, this one leaves us unsure as to whether two processes are operating or just one. That is, to consider Group A+/B+, the test result could be a consequence of heightened generalization between B and A, of restriction in generalization between C and A, or both of these. We are obliged, therefore, to offer a theoretical analysis for both possibilities.

**Analysis**

What determines the degree of generalization between two stimuli, A and B? Figure 4 offers one representation of the standard interpretation. Each stimulus can be construed as comprising a set of elements or features, some of which it will hold in common with other stimuli (two auditory events might, for instance, come from a common source, have the same intensity and duration, etc.) and some of which will be unique (the events might be tones differing in frequency). In Figure 4, the elements that make up Stimulus A are represented by a circle, unique features of A are shown by the section *a*, and features held in common (*c*) with Stimulus B are shown by



**Figure 3. Group mean suppression ratios for test trials with Stimuli B and C from the experiment by Honey and Hall (1989). All subjects had experienced A, B, and C in Stage 1 training. For Group A+/B+, A and B had both signaled food; for Group A-/B+, only B had signaled food in Stage 1. In Stage 2, A was used to signal shock.**



**Figure 4.** Stimuli A, B, X, and Y are represented by circles; arrows represent associative links. Each stimulus is thought of as a set of features or elements, some of which are held in common with those of other stimuli. The overlap between circles indicates the proportion of common elements shared. Thus, A and B have unique features (*a* and *b*) and a set of common elements (*c*).

the area of overlap with the circle representing B. Conditioning with A as a CS will endow all its features with associative strength, and the strength governed by the *c* elements will mean that the untrained B will also be capable of evoking a certain amount of (generalized) responding. The task of our analysis is to explain why equivalence/distinctiveness training should modify the extent of such generalization.

The upper panel of Figure 4 indicates the associations that will form as a result of equivalence training in which A and B are both associated with the common consequence X. Introducing the assumption that an associatively activated representation is capable of acquiring associative strength permits this associative structure to predict enhanced generalization between A and B. Subsequent reinforced training with A and some US will establish a link between the associatively activated X representation and the US representation. Stimulus B would then be able to activate the US representation (and thus elicit conditioned responding) by virtue of its ability to activate X. Generalized responding would thus be mediated not only by way of the intrinsic common elements (*c*) but also by way of a set of elements (X) that the stim-

uli have come to share by way of their common training history. The notion that the image of an event can come to act as a CS thus supplies a ready explanation for the phenomenon of acquired equivalence.

The explanation of acquired distinctiveness is a little more problematic. The lower panel of Figure 4 represents the outcome of discrimination training in which the target cues, A and B, have become associated with rather different consequences, X and Y (they are shown in Figure 4 as having no stimulus elements in common). It is not immediately obvious that this pattern of associative links would reduce generalization between A and B. The existence of the A–X association would, according to our assumptions, allow the representation of X to gain strength during reinforced training with A; however, as long as the *c* elements gain strength, generalization to B will occur, and discrimination will not be improved. Acquired distinctiveness requires that, in some way, the role of these *c* elements be reduced. There is one possible way in which this might be achieved (Hall, 1991, discusses others) that follows directly from the assumptions we have made so far. We have assumed that an associatively activated representation can acquire associative strength through pairing with a US. If it acts like any other CS, then we must expect the rules governing acquisition to be the same as those that govern learning about directly activated representations. In particular, overshadowing can be expected (see Holland, 1983). The acquisition of strength by X during reinforced training with A should, to some extent, overshadow A, restricting the strength gained by all its features. It will thus limit the acquisition of strength by *c* elements, generalization between A and B will be limited, and discrimination will be enhanced.<sup>4</sup> In summary, then, the notion that the image of an event can function as a CS can provide a coherent account of the basic phenomena of both acquired equivalence and acquired distinctiveness. We now consider the application of this analysis to some less familiar demonstrations of these effects.

### Extension to Other Conditioning Procedures

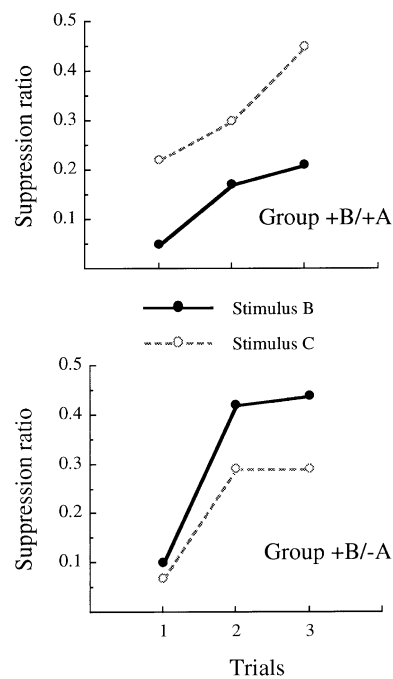
The essence of acquired equivalence is that generalization between stimuli is enhanced as a result of their having experienced a common training history. This common training typically involves their both having formed an association with a given consequence. But there are, of course, other ways in which events can share a history. One of these was explored in an experiment by Honey (1990), the design of which is summarized in Table 5b.

This study used rats as the subjects and assessed generalization from one auditory cue (A, food-reinforced in Stage 2) to two others (B and C). Cues A and B shared a common training history in that, in Stage 1, both had been repeatedly presented, nonreinforced. The final test showed that the CR established in Stage 2 generalized more readily to B than to the novel C. Honey (1990) interpreted this result in terms of a version of the acquired equivalence mechanism described above. Stage 1 train-

ing, he argued, will endow A and B with common properties. Both, for instance, might become linked with some central representation encoding that no event follows a stimulus presentation. Activation of this associate during Stage 2 conditioning will allow it to acquire associative strength; its activation by B during the test will allow this stimulus to evoke more conditioned responding than the untrained Stimulus C. Such a mechanism could play a role in the effect obtained by Honey and Hall (1989) in their Group A-/B+ (Table 5a). That there was less generalization from A to B than from A to C in that experiment can be explained solely in terms of acquired distinctiveness (A and B received different treatment in Stage 1); but A and C were treated in the same way in Stage 1 (both were presented nonreinforced), and acquired equivalence mediated by their common association with no event could therefore have contributed to the extra generalized responding controlled by C in the test stage. Perhaps, however, in this experiment, the mediating representation was more likely to be of "no US" rather than simply of "no event," given that the reinforced training given to B in Stage 1 might have resulted in inhibitory learning to A and C. Further evidence relevant to the possibility of mediation by way of inhibition comes from the experiment to be considered next.

Hall et al. (1993) reported a series of experiments intended to demonstrate what they referred to as *backward equivalence*. The design of one of their experiments is shown in Table 5c. It will be evident that this study was identical to that conducted by Honey and Hall (1989) except that, in Stage 1, the delivery of food occurred before the presentation of the relevant auditory cue. The results (Figure 5) were identical—that is, generalization of suppression from Stimulus A occurred more readily to the test stimulus with which it had shared a common Stage 1 training history. Hall et al. concluded that equivalence would be established by training in which the stimuli share a common antecedent as well as by the more usual procedure in which they are trained with a common consequence. Hall et al. considered several possible mechanisms that might be responsible for backward equivalence, one of which was that inhibitory learning might mediate generalization. Consider Group +B/+A (Table 5c). There is reason to think (e.g., Wagner & Larew, 1985) that the backward pairings experienced in Stage 1 might establish both A and B as inhibitory CSs, each of them acquiring the ability to evoke a given central state (e.g., frustration engendered by activation of a no-US representation). Aversive conditioning with A in Stage 2 would then establish frustration as a signal for shock and thus allow B, which also evokes this state, to elicit the CR in the test stage. This analysis is the exact parallel of that applied to *forward equivalence* (Honey & Hall, 1989) but with activation of a no-US representation rather than activation of a US representation serving as the mediating event.

An alternative (or additional) possibility is that the backward equivalence training procedure brings into play a set of associative processes that result in the formation of a direct link between the relevant stimuli



**Figure 5.** Group mean suppression ratios for test trials with Stimuli B and C from the Experiment by Hall et al. (1993). All subjects had experienced A, B, and C in Stage 1 training. For Group +B/+A, A and B had both been preceded by food; for Group +B/-A, only B had been preceded by food in Stage 1. In Stage 2, A was used to signal shock.

(in this case, A and B). In the procedure used here for Group +B/+A, food is made available shortly before each auditory cue; the presentation of food, therefore, will become a CS associated with both B and A. As a result, the representation of B would be activated on +A trials and the representation of A would be activated on +B trials—that is, each auditory cue would occur along with the associatively activated representation of the other. Associations might then be formed between A and B during Stage 1 training, in spite of the fact that the two events themselves had not occurred together. A CR established to A could then be evoked by B by way of an associative chain B-A-US, such as that taken to be responsible for some instances of sensory preconditioning (see above; see also Honey & Hall, 1991).

Evidence to support the reality of this second alternative comes from an experiment by Bonardi, Rey, Richmond, and Hall (1993), which demonstrates backward equivalence using a procedure in which inhibitory learning seems less likely to play a role. The design is outlined in Table 5d. Pigeons were autoshaped in Stage 1 with four keylight stimuli (A, B, C, and D), presentation of each of which was followed by food. On the basis of the evidence and arguments presented so far, this training procedure might be expected to enhance generalization among these stimuli, with food representation acting as the mediating event. We hoped, nonetheless, that another feature of the Stage 1 training would permit a backward equivalence effect to emerge. The four trial

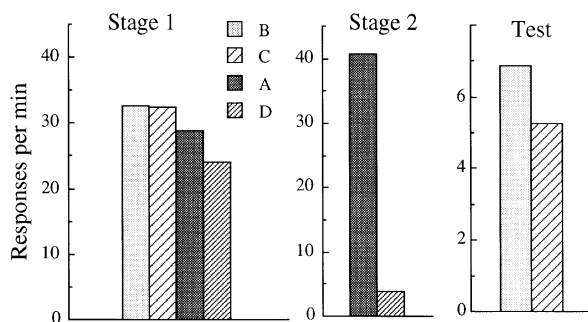
types were divided into two pairs: Presentations of A and C were always preceded by Stimulus X, and presentations of B and D were always preceded by Stimulus Y. Backward equivalence would be demonstrated if generalization could be shown to be particularly marked between the stimulus pairs that shared a common antecedent in this stage. In Stage 2, one member of each pair received further training. Continued reinforcement increased the level of responding controlled by A; nonreinforcement led to a loss of responding to D. The final test (Figure 6) showed that the effects of this training generalized differentially to B and C. Response levels were low (the test was carried out in extinction), but Stimulus B evoked significantly more responding than did Stimulus C. That is, the training procedure was effective in establishing equivalence between A and C and/or between B and D.

In the instances considered so far, equivalence between stimuli has been established in one of two ways: by training in which they have signaled a common event (or the absence of an event in the experiment by Honey, 1990) and by training in which they have themselves been signaled by a common event. That is, the critical stimuli have been trained either as CSs or as USs. Are there other properties that stimuli can acquire as a result of training that might serve to establish equivalence between them? In fact, such is the power of associative theory (or, perhaps, such is the narrow-mindedness of learning theorists) that almost all of our standard training procedures have been interpreted in terms of the formation of associations in which the only function of stimuli is to act either as signals (CSs) or as the targets of signals (USs). One possible exception is occasion setting. The ability of animals to learn a discrimination in which a given cue, A, is reinforced in the presence of X but nonreinforced in the absence of X [ $X(A+)/A-$ ] seems to require, for its explanation, learning mechanisms that go beyond the formation of simple CS-US associations (e.g., Swartzentruber, 1995). It is interesting to ask, therefore, whether equivalence will be established (generalization will be enhanced)

between cues (e.g., X) that have served as occasion setters and between cues (such as A in the above example) that have been subject to the action of occasion setters. Bonardi and Hall (1994) attempted to answer the second of these questions.

The essential features of this experiment are summarized in Table 5e. Rats were first trained on two conditional discriminations: A was reinforced only in the presence of Occasion Setter X; B was reinforced only in the presence of Occasion Setter Y. The animals came to show the CR to A and B when these stimuli occurred in the presence of their occasion setters but learned not to respond to A or B presented alone. In Stage 2, Stimulus A received continuous reinforcement so that a vigorous CR was now established. The test stage showed that this CR generalized readily to Stimulus B but not to Control Stimulus C, which, like A and B, had been subjected previously to both reinforcement and nonreinforcement but which had not been trained in an occasion-setting task. Bonardi and Hall (1994) concluded that training A and B as the targets of occasion setters had been effective in establishing equivalence between them.

Holland (e.g., 1989), on the basis of his studies of transfer effects in occasion setting, has come to the conclusion that occasion setting involves a memory system different from that concerned with simple associations. One interpretation of our results, therefore, might be that A and B are rendered equivalent by virtue of the fact that both have representations of some sort in this system. Proper evaluation of this suggestion must await a more detailed specification of the properties of this special memory system. We can, however, pursue straight away the possibility that our results might be explained without recourse to any special assumptions but in terms of the associative processes that have held good so far. Bonardi and Hall (1994) point out that the essence of the associative account of (forward) equivalence is that training two stimuli as signals for a common consequence effectively adds a new set of (associatively activated) common elements to each. They ask, then, whether the occasion-setting training shown in Table 5e can be construed as endowing A and B with extra common elements in this way, and they conclude that associations between the occasion setters and the target CSs that occur in their presence might play this role. Although the stimuli used as occasion setters by Bonardi and Hall were nominally quite different (one was an increase in general illumination, the other was a decrease), they nonetheless had several common characteristics (e.g., both were visual events of the same relatively long duration). And although the occasion-setting procedure may involve the operation of learning mechanisms that lie outside the scope of our usual associative principles, there is no reason to think that this excludes the formation of associations between the target CSs and the occasion setters. Accordingly, A and B should acquire the ability to activate the representations of Occasion Setters X and Y, and the associative strength acquired by the elements common to X and Y during rein-



**Figure 6.** Group mean performance in the experiment by Bonardi et al. (1993, Experiment 2). A, B, C, and D represent four different keylight stimuli. Scores for Stages 1 and 2 are for the final session of each stage; test scores are pooled over all test sessions. The design of the experiment is summarized in Table 4d.

forced Stage 2 training with A could thus allow B to evoke the CR on the test.

### Extension to Equivalence Classes in Discrimination Learning

Recent years (since the pioneering work of Sidman, 1971) have seen the growth of an extensive literature analyzing the way in which certain forms of discrimination training can lead to the formation of *equivalence classes*. Two stimuli may be said to be equivalent in some respect when one can function as a substitute for the other in that respect; it is in this sense that we have used the term so far. Sidman (1990) refers to such stimulus substitutability as *functional* equivalence and distinguishes it from the formal notion of equivalence that is his central concern; he adopts the logical or mathematical definition of equivalence as describing a relationship between events that shows the properties of reflexivity, symmetry, and transitivity. Such equivalence can emerge, it is suggested, through training on some types of discrimination task (sometimes called *symbolic matching-to-sample tasks*), and appropriate variations of these tasks can be used to reveal the existence of the three defining properties of equivalence. Our concern in what follows is to determine whether the associative processes that we take to underlie functional equivalence might also be operating in the symbolic matching procedure and, if they are, whether they can help supply an explanation for the effects that are obtained.

A simple version of the conditional discrimination procedure usually employed in studies of equivalence class formation is presented in Table 6a. In an experiment with pigeons as subjects, Sample Stimulus  $A_1$  might be, for instance, a red keylight; its offset would be followed by two comparison stimuli,  $B_1$  and  $B_2$  (say horizontal and vertical), and a peck to  $B_1$  would be reinforced. On other trials, Sample Stimulus  $A_2$  might be a green keylight, and response to the other comparison stimulus,  $B_2$ , would be reinforced. In an experiment with young children as the subjects, Sample Stimuli  $A_1$  and  $A_2$  could be the spoken words *dog* and *cat*, Comparison Stimuli  $B_1$  and  $B_2$  would be pictures of these animals, and appropriate choice might be rewarded by approval. This task is formally identical to that described for pigeons, and, for very young children, the relationship among the stimuli will be just as arbitrary.

The solution of a discrimination of this type requires the involvement of some sort of conditional learning mechanism. One possible analysis views  $A_1$  as an occasion setter that selectively activates the  $B_1$ -US association, with  $A_2$  having similar control over the  $B_2$ -US association. But, as we have already observed, the occurrence of conditional learning does not preclude association formation, and, in this procedure, it is likely that an association will be formed between the sample stimulus and the rewarded comparison stimulus that follows it: As the discrimination starts to be learned, the subjects will, by virtue of their own responding, come to experience these two events regularly and in close succession. Thus,  $A_1$  will acquire the power to activate the representation

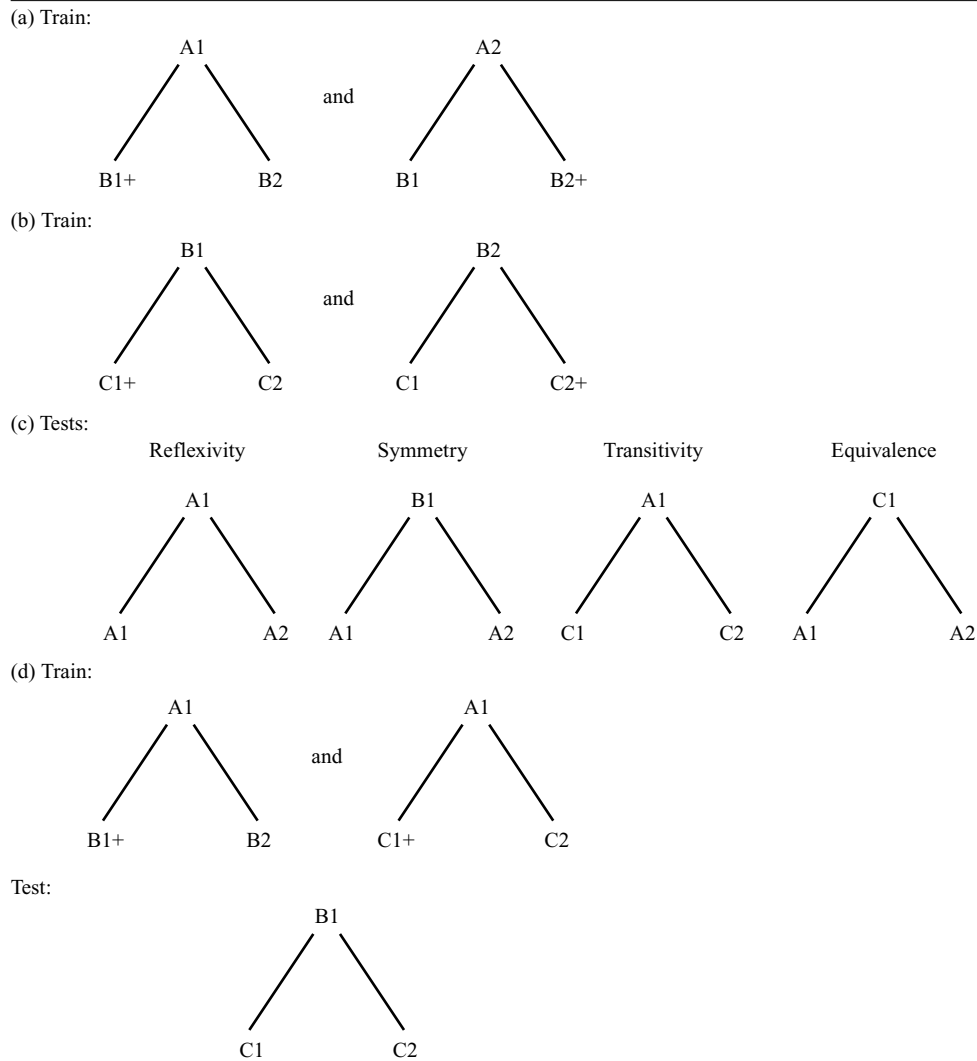
of  $B_1$ , and  $A_2$  will be able to activate  $B_2$ . Adding our basic assumption that associatively activated representations will support learning and performance in much the same way as directly activated representations allows explanation of the effects seen in some of the transfer tests used in studies of equivalence class formation.

Table 6c represents four of the various tests that are employed in these experiments. The first two (for reflexivity and symmetry) require initial training only on the task shown in Table 6a; the tests for transitivity and for equivalence itself require, in addition, training on a further conditional discrimination, with the comparison stimuli from the first task now serving as samples for a new set of comparison stimuli (Table 6b). Let us consider, for the time being, just the test for transitivity. Here, the subject is given a choice of comparison stimuli ( $C_1$  and  $C_2$ ) following presentation of a sample ( $A_1$ ) with which they have not received training. Choice of  $C_1$  is taken to indicate that the tasks on which they have been training have established equivalences (that can be thought of as  $A_1 = B_1$ , and  $B_1 = C_1$ ), allowing the transitive inference  $A_1 = C_1$ . After some early failures (e.g., D'Amato, Salmon, Loukas, & Tomie, 1985), appropriate responding on the transitivity test has now been successfully demonstrated in nonhuman subjects (Kuno, Kitadate, & Iwamoto, 1994).

The associative explanation of this result follows directly from the proposition that training on the first discrimination will allow  $A_1$  to come to activate the representation of  $B_1$ . This means that  $B_1$  will be activated on presentation of the  $A_1$  sample in the test.  $A_1$  itself will not be able to direct choice between  $C_1$  and  $C_2$ , but  $B_1$ , having been explicitly trained in this role in the second discrimination (Table 6b) should be able to do so. That is,  $A_1$  will be an effective sample because it evokes the image of an event that has received appropriate training. For such a system to work requires that the  $A_1$ - $B_1$  link be well formed initially, and in this may lie the reason for the success of Kuno et al. (1994). They gave extensive discrimination training with relatively prolonged exposure to both sample and comparison stimuli (their subjects, pigeons, were required to peck 10 times at each stimulus on each trial). These are conditions that would foster the formation of a strong association between the sample and the correct comparison stimulus.

Similar arguments can be applied to the analysis of the experimental procedure outlined in Table 6d. This is based on the original study by Sidman (1971) and is of special interest and importance because it is described by him (Sidman, 1990) as being the "basic experiment" in the demonstration of equivalence relations. Here, the subject is trained on two conditional discriminations with a given sample stimulus:  $A_1$  indicates that  $B_1$  should be chosen rather than  $B_2$ , and, on other trials,  $C_1$  should be chosen rather than  $C_2$ . (The companion tasks in which  $A_2$  is used to signal that  $B_2$  and  $C_2$  should be chosen are not shown in the table.) In the test, one of the original comparison stimuli is used as the sample; in the example given, the question posed is, Will using  $B_1$  as the sample allow the subject to choose between  $C_1$  and  $C_2$ ? To put

**Table 6**  
**Conditional Discrimination Procedures and Tests for Equivalence**



Note—Letters represent stimuli. In each triad, the upper letter represents a sample stimulus, the lower pair represents comparison stimuli. The + indicates the comparison stimulus that is to be chosen in training.

flesh on these bones:  $A_1$  might be the spoken word *dog*,  $B_1$  a picture of a dog, and  $C_1$  the written word *DOG*. Sidman (1971) found, for his nonverbal subject, that training in which the spoken word came to control choice both of the picture and of the written word allowed appropriate choice on the test; that is, the picture as sample now controlled choice of the written word.

A possible associative interpretation of this result is as follows. Training on the initial discriminations establishes associations that allow Stimulus  $A_1$  to activate the representations of  $B_1$  and  $C_1$ . This means, for instance, that  $B_1$  will be activated on those training trials on which  $A_1$  is gaining control over the choice of  $C_1$ . The  $B_1$  representation will thus acquire similar control, with the result that when  $B_1$  is actually presented as a sample in the final test, the appropriate choice response will be evoked.

Whether this possible explanation is in fact the correct explanation is another matter. If associative processes do indeed underlie equivalence effects of this sort, then such effects might be expected in any organism capable of associative learning. The evidence on this matter is inconclusive. The closest parallel in studies using animal subjects is found in experiments on “one-to-many” matching in pigeons by Urcuioli and Zentall (1993; see also Urcuioli, Zentall, & DeMarse, 1995). In these experiments, the birds received training on versions of the tasks shown in Table 6d, and then equivalence between  $B_1$  and  $C_1$  was assessed. This was achieved by training the  $B_1/B_2$  discrimination with a new sample stimulus followed by a test in which the ability of this new sample to control choice between  $C_1$  and  $C_2$  was evaluated. Evidence of appropriate choice was slight (Urcuioli & Zentall, 1993)

or nonexistent (Urcuioli et al., 1995). It should be noted, however, that this test procedure is substantially more complex than that of the Sidman design (Table 6d), and successful performance would require a correspondingly more complex chain of associations. To this extent, the associative explanation of Sidman's "basic experiment" remains viable.

Although associative processes can generate an explanation for transitivity and for the equivalence effect of Table 6d, this does not mean that they can explain equivalence phenomena that possess all three of Sidman's (1990) defining features: reflexivity, symmetry, and transitivity (Table 6c). The point is that the sort of equivalence demonstrated by the "basic experiment" of Table 6d does not require that all these features be operating. To revert to the notation used previously, the subject in this experiment need learn only  $A = B$  and  $A = C$ ; those mechanisms that we have used previously to provide an explanation for transitivity will then allow correct performance ( $B = C$ ) on the test. The test does not involve the relation of symmetry (i.e.,  $A = B$ , therefore  $B = A$ ). Symmetry is, however, required for correct performance on another of the tests of equivalence used in experiments of this sort—that shown in Table 6c, in which the standard test for transitivity is reversed and an event previously trained as a comparison stimulus appears for the first time as a sample. Studies with human subjects have shown them to be capable of appropriate performance on this test (e.g., Sidman & Tailby, 1982; Stromer & Osborne, 1982). There is, however, no uncontested demonstration of this result in nonhuman animals (see Hayes, 1989; Saunders, 1989).

If the transfer shown in these discrimination tasks depends on associative processes of the sort we have described, then it is not entirely surprising that animals should fail the full equivalence test of Table 6c. Associative links formed in the first stage of training (Table 6a) might allow  $A_1$  and  $A_2$ , when presented as comparison stimuli in the test, to evoke representations of  $B_1$  and  $B_2$ . But the new sample stimulus ( $C_1$ ) would be able to evoke the representation of the trained sample ( $A_1$  for the choice between  $B_1$  and  $B_2$ ) only by way of a chain of backward associations (i.e.,  $C-B-A$ ). Such backward associations are not readily formed. Animals trained to choose  $B_1$  after presentation of  $A_1$  do not tend to choose  $A_1$  when the relationship is reversed and  $B_1$  is now used as the sample (D'Amato et al., 1985; Hogan & Zentall, 1977). In different terminology, these training procedures do not establish the symmetry relation, and, hence, the equivalence test—which depends on the effectiveness of this relation—will be failed.

It may be tempting, on the basis of this analysis, to conclude that humans and animals differ in their transfer performance on these conditional discriminations because they have different learning mechanisms available to them: Animals rely on associative processes that allow them to show transitivity but do not allow them to pass the full equivalence test; people, who can succeed on the equivalence test, must (it may be thought) be making use of some other learning process that the animals lack. It

would, however, be premature to accept this conclusion before examining the rather simpler alternative that the difference between people and animals in their performance on these tests has its origin in differences in the procedure employed in the two cases. Although formally equivalent, the procedures used for different species are necessarily somewhat different in detail. We may ask whether these details make the test unfair to animals, preventing them, in some way, from showing that they possess the full range of learning processes available to people; alternatively, they may be too fair to people, allowing them to employ associative processes when animals cannot.

Zentall et al. (1992) note a number of features of the procedure used in studies of animals that might tend to hinder the ability of these subjects to learn the equivalence relation. One point they do not make, but which may be of particular importance in the context of the analysis presented here, concerns the timing of presentation of the sample and comparison stimuli. In experiments with animals, the standard procedure permits the formation of an  $A_1-B_1$  association, with  $A_1$  becoming a CS for  $B_1$ ; but, as has already been noted, the backward association ( $B_1-A_1$ ), which would allow  $B_1$  to act as a CS capable of exciting the representation of  $A_1$ , is unlikely to be formed. By contrast, in studies with human subjects, it is common practice for the sample stimulus to remain on when the comparison stimuli are presented. Simultaneous presentation of these events can be expected to generate both  $A_1 \rightarrow B_1$  and  $B_1 \rightarrow A_1$  associations. These associations could be responsible for correct performance on the equivalence test shown in Table 6c. Although  $A_1$  and  $A_2$  have not been trained as comparison stimuli, they might activate associates ( $B_1$  and  $B_2$ ) that have. And although  $C_1$  has not been trained as a sample with  $B_1$  and  $B_2$  as comparison stimuli, it would activate the representation of a stimulus ( $A_1$ ) that has been trained in this way by way of the chain  $C_1-B_1-A_1$ .

As it stands, the analysis just presented is best viewed as an exercise in determining whether or not our associative analysis could apply, in principle, to the outcome of conditional discriminations designed to assess the formation of equivalence classes. Whether the analysis should be accepted as correct will depend on the outcome of further experiments. We may conclude, however, that it would be unwise to assert, at this stage, that the phenomena revealed in experiments on equivalence classes necessarily imply the operation of processes beyond those assumed by standard theories of conditioning.

## PERCEPTUAL LEARNING

Nonreinforced preexposure to a pair of stimuli can enhance the ease with which animals subsequently learn a discrimination between them. The first part of this section of the article gives some concrete examples of this perceptual learning effect. There follows an analysis of these procedures in terms of the role that image conditioning is likely to play in them. This reveals a problem



in that these theoretical considerations appear to predict the opposite of the results obtained. The section concludes, therefore, with an attempt to show how this problem might be resolved.

### The Phenomenon

The classic example of the perceptual learning effect is the demonstration by E. J. Gibson and Walk (1956) that rats are better able to learn a simultaneous discrimination between geometrical figures after prolonged preexposure to these figures presented in their home cages (see also Hall, 1979, 1980). But the effect has been demonstrated in a wide range of training procedures. Chamizo and Mackintosh (1989; see also Trobalon, Sansa, Chamizo, & Mackintosh, 1991) have found that separate non-reinforced exposure to each of the relevant maze arms can facilitate subsequent discrimination learning in which rats are rewarded for choosing one arm rather than the other. Studies of the domestic chick (e.g., Honey, Horn, & Bateson, 1993) have shown that discrimination between two cues is enhanced by prior training designed to allow imprinting to each of the cues. In studies of flavor-aversion learning, it has been shown that an aversion conditioned to one flavor will generalize less well to another when the animals have had prior experience of these flavors.

A study by Symonds and Hall (1995, Experiment 1) provides a convenient example of the perceptual learning effect in flavor aversion. The stimuli were two flavors (saline and sucrose solutions) rendered rather similar (it may be assumed) by the addition of the sour taste of acid to each of them. Discrimination was assessed by establishing an aversion to one of them (Stimulus A was presented on three occasions prior to an injection of 0.3 M LiCl) and by measuring the extent of generalization of this aversion to the other (B) on a subsequent test trial. The results of this test are shown in Figure 7. The four groups shown in the figure differed in their previous experience of the flavors. Group A/B had received four preexposure presentations of A intermixed with four of B; Group A received four preexposures to Flavor A; Group B received four preexposures to Flavor B; Group W received only unflavored water during the preexposure phase. The low levels of consumption shown on the test by all of the last three groups indicate that the aversion acquired by A generalized to B in all of them. Group A/B showed much less generalization: Preexposure to both A and B appeared to enhance the ability of the subjects to discriminate between them.

### Analysis

The existence of the perceptual learning effect constitutes something of a problem for the theoretical analysis developed in previous sections of this article. It appears to be an instance of the acquired distinctiveness of cues, but it arises from a training procedure that lacks what has been taken to be critical in producing the effect—that is, the preexposure phase gives no opportunity for the cues to become associated with distinctively different events. Indeed, given that the critical stimuli receive just the

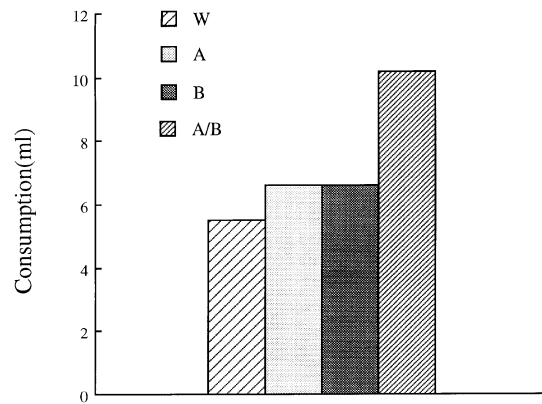
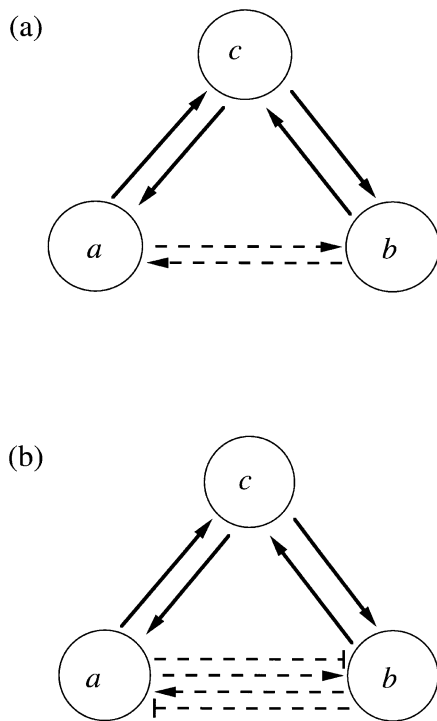


Figure 7. Group mean scores on the test session with Flavor B after conditioning to A in the experiment by Symonds and Hall (1995, Experiment 1). Group A had received preexposure to A, Group B had received preexposure to the Flavor B, Group A/B had received preexposure to both flavors, and Group W had received only water in preexposure.

same treatment during preexposure, the outcome of this form of training might be expected to be acquired equivalence rather than acquired distinctiveness. The experiment by Honey (1990), described above, provides direct evidence that nonreinforced preexposure to a pair of stimuli can enhance generalization between them. Admittedly, Honey's (1990) experiment made use of stimuli and procedures rather different from those used in most studies of perceptual learning, but this observation allows no solution to the problem. In fact, the nature of the stimuli used in experiments on perceptual learning is such as to increase the likelihood of associatively mediated generalization occurring between them, as the following will show.

If preexposure is to produce a perceptual learning effect, it is necessary to use stimuli that are fairly similar to start with; if the stimuli were perfectly discriminable from the outset, there would be no scope for preexposure to produce an enhancement in discriminability. In terms of the representational scheme used in Figure 4, the overlap between the critical stimuli, A and B, needs to be extensive, with many elements ( $c$  elements) being held in common. Indeed, in some experiments,  $c$  elements are explicitly introduced in order to ensure this, as in the experiment by Symonds and Hall (1995) in which acid was added to both the sucrose and the saline solutions. The acid presumably increases the number or salience of the  $c$  elements and thus ensures that generalization will occur readily between A and B. Figure 8a presents the three critical components of the stimuli in a slightly different way. Here  $a$  and  $b$  represent nodes sensitive to the unique features of A and B (in this case, salt and sweet), and  $c$  represents a node sensitive to common elements (sour plus any other features that saline and sucrose solutions intrinsically hold in common). Preexposure to A and B will allow associations to form among these nodes. Presentations of A will activate  $a$  and  $c$  concurrently so that the excitatory  $a-c$  and  $c-a$  associations (shown by solid lines in the figure) will form. Similarly, presenta-



**Figure 8.** Circles represent the unique features (*a* and *b*) of two Stimuli A and B and also the features they hold in common (*c*). Lines connecting the circles indicate associations that might be formed among them. Solid lines represent standard excitatory links; dashed lines represent links that may be formed when one of the elements is activated associatively. A stopped end indicates an inhibitory link.

tions of B will result in associations between *b* and *c*. Once these links have been formed, another set of associations will be able to develop. Presentation of Stimulus A will, by way of the *c*–*b* link, be able to induce activity in the *b* node; similarly, presentation of B will be able to induce activity in the *a* node. Thus, on each of these trials, the two nodes representing unique stimulus elements will be activated concurrently, one directly and one by way of an associative link. Our hypothesis has been that associations will form in these conditions; these (shown by the dotted lines in the figure) will allow each set of unique elements to induce activity in the other set.

The existence of these various associations will tend to enhance generalization between A and B. For animals trained on a discrimination of A+/B– without any prior exposure to the stimuli, the only source of generalization between A and B (if we neglect the possibility that within-stimulus associations might form during the course of discrimination training itself) will be the associative strength directly controlled by the *c* elements. But for animals given preexposure to A and B, there are two further possible sources of generalization: (1) Reinforced A trials will allow the associatively activated representation of *b* to become linked to the US, thus increasing the like-

lihood that B will tend to evoke the CR. (2) On test trials with Stimulus B, the associative connections shown in the figure will result in the activation of *a*, a set of elements directly paired with the US during conditioning, and thus likely to be powerfully able to evoke the CR. That is, the associative structure shown in Figure 8a predicts quite the reverse of the perceptual learning effect.

An experiment by Honey and Hall (1991) serves to demonstrate the reality of mediated generalization produced by within-compound associations. Rats in the experimental condition were given exposure to two flavors, A and B, each mixed with a third to form the compound stimuli AX and BX. Control subjects also received exposure to two compounds, but these were AY and BY (i.e., A and B were mixed with different flavors). Both groups then received training designed to establish a conditioned aversion to A, and, finally, generalization to B was tested. The experimental group was much less willing to consume Flavor B in the test than was the control group—that is, prior experience of A and B, along with a salient common element (X), enhanced generalization between them. Honey and Hall (1991) presented this result as an instance of acquired equivalence and interpreted it in terms of their standard (e.g., Honey & Hall, 1989) explanation for this effect, suggesting that an A–X association allows the image of X to acquire strength on the reinforced A trials and that a B–X association allows the image of X to influence performance on the test trials. What is more, in this experiment, the use of simultaneous compounds in preexposure could allow the formation of links between the unique features of A and B, as shown in Figure 8a. These links, too, will enhance generalization by allowing the representation of B to be reinforced on conditioning trials and the representation of A (the CS) to be evoked on the test trial.

### Possible Solutions

The theoretical arguments presented so far lead to the conclusion that there should be no perceptual learning effect; rather, preexposure to stimuli should enhance generalization between them. In order to accommodate the facts, therefore, it is necessary to modify or extend the theory. It may be, for instance, that the analysis presented so far is merely incomplete—that the preexposure procedures used in studies of perceptual learning do indeed tend to establish equivalence between the stimuli but that, in these procedures, some other process that tends to reduce generalization is also at work. In perceptual learning experiments, it must be assumed, the effects of the latter process outweigh those of the former. What we must determine now, therefore, is what this extra process might be and why it should be especially powerful in procedures that generate a perceptual learning effect. We will consider two possibilities, neither of which requires a radical departure from the set of associative learning principles that have served well to this point.

**Latent inhibition.** Preexposure to the target stimuli in a perceptual learning experiment will result in latent inhibition, restricting the readiness with which these stimuli

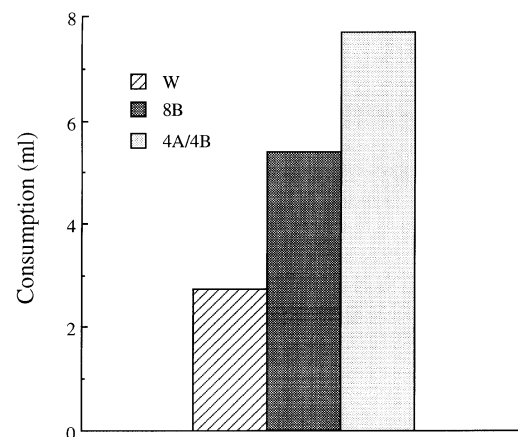
enter into associations when reinforcement is subsequently introduced. In terms of the scheme used in Figure 8, preexposure to A ( $ac$ ) and to B ( $bc$ ) will produce latent inhibition to all elements of the stimuli ( $a$ ,  $b$ , and  $c$ ). But, as McLaren, Kaye, and Mackintosh (1989) have noted, this preexposure will be particularly effective in establishing latent inhibition to the  $c$  elements, since these (unlike  $a$  and  $b$  elements) will be present on all preexposure trials. Now, the stimuli used in studies of perceptual learning are typically similar to each other and generalization occurs readily between them; that is, they have particularly potent or numerous  $c$  elements, and the associative strength governed by these elements limits performance on discrimination between them. Elimination (or a dramatic attenuation) of the effectiveness of the  $c$  elements as a consequence of preexposure to the stimuli would thus restrict generalization between A and B and could thus produce a perceptual learning effect. Evidence to favor this account comes from a study of maze learning by Trobalon, Chamizo, and Mackintosh (1992) in which the perceptual learning effect was found to be abolished by a change of context between the preexposure and discrimination stages. It is well established that a context change will disrupt latent inhibition (e.g., Channell & Hall, 1983), and, thus, such a change can also be expected to disrupt a perceptual learning effect in which latent inhibition plays an important role.

The results shown in Figure 7 could also be explained in these terms. In that experiment, Group A received four preexposures to Flavor A, and Group B received four preexposures to Flavor B. Thus, each of these groups received four presentations of  $c$ . If four preexposures produce little or no latent inhibition, then these groups will condition as well to  $c$  as will Group W, and the generalization test performance of these three groups will be similar. But Group A/B received eight presentations of  $c$  (four of A and four of B). If this is enough to produce substantial latent inhibition to  $c$ , then the  $c$  elements will acquire little associative strength during conditioning with A, and Flavor B will be consumed readily on test—the result obtained. However, there is reason to think that, in this training preparation, latent inhibition plays no more than a minor role. An experiment by Symonds and Hall (1996) demonstrates that their perceptual learning effect is not to be explained (solely) in terms of differences between groups in the latent inhibition suffered by  $c$  elements. This experiment (see also Mackintosh, Kaye, & Bennett, 1991) included Groups A/B and W, whose treatment exactly matched that of the groups run by Symonds and Hall (1995, Experiment 1). A third group (Group B) differed only in that preexposure consisted of eight presentations of B. All subjects then received conditioning to A followed by a generalization test with B. The results (Figure 9) confirm the previous finding in that Group A/B showed less generalization from A to B than did the nonpreexposed group (Group W). However, Group B, although it showed less generalization than Group W, showed substantially and significantly more generalization than Group A/B. [An ANOVA conducted

on the data summarized in the figure showed there to be a significant difference among the groups,  $F(2,21) = 16.38$ ,  $p < .01$ . Pairwise comparisons showed that each group differed from each of the others,  $ps < .05$ .] The critical point here is that Groups B and A/B will not differ in the latent inhibition that has accrued to the  $c$  elements—both have had eight preexposures to these elements. Some other process must be operating in Group A/B to produce the additional restriction in generalization that prior exposure to both A and B reveals.

**Inhibition between unique elements.** The phenomenon of perceptual learning presents a problem for our associative analysis because the presence of salient  $c$  elements in the stimuli allows the formation of excitatory associations among various components of the stimuli, associations that will act to enhance generalization. But this problem may arise not because this analysis is flawed but rather because it has not been carried far enough. McLaren et al. (1989) have argued that the formation of the excitatory links shown in Figure 8a is only the first stage of the learning that goes on during preexposure. Once these links have been formed, further presentations of A will activate the representation of the unique features of B, and this will happen in the absence of B itself. Similarly, further presentations of B will activate the  $a$  representation in the absence of A. These are circumstances (the omission of an expected event) that are commonly thought to foster inhibitory learning and should, according to McLaren et al. (1989), lead to the formation of the inhibitory links between  $a$  and  $b$  that are shown in Figure 8b. To the extent that generalization from A to B occurs because B is able to activate the representation of  $a$ , subjects given prolonged preexposure can be expected to show less generalization; for them, presentation of B will inhibit the activation of  $a$ .

Evidence consistent with this interpretation comes from an experiment by Symonds and Hall (1995, Exper-

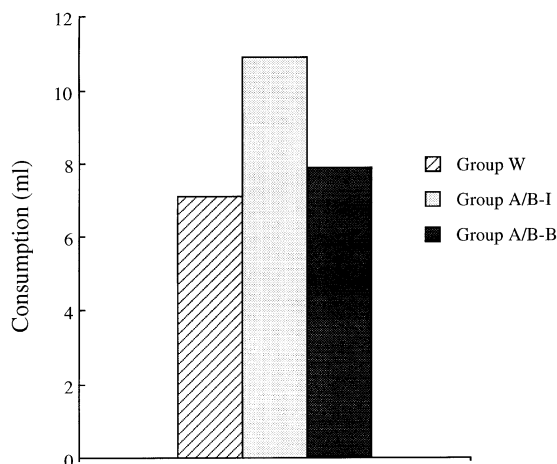


**Figure 9.** Group mean scores on the test session with Flavor B after conditioning to A. Prior to conditioning, Group 4A/4B received preexposure in which each flavor was presented four times; Group 8B received all trials with B; Group W received only water in preexposure.

iment 2). This used the same flavor-aversion learning procedure as was described earlier; indeed, two of the conditions (Group W and Group A/B-I) were identical to those studied previously, with Group W receiving no preexposure to the flavors and Group A/B-I (like Group A/B of the previous experiment) receiving exposure to both A and B. The suffix I indicates that trials with A and B were intermixed, presentations of one stimulus being given each morning and presentations of the other given 6 h later in the afternoon, throughout the preexposure phase. A third condition (Group A/B-B) was novel. Here, the suffix B indicates that presentations of the stimuli were "blocked"; rather than experiencing both flavors on the same day, these subjects received a block of trials, with one of them followed by a block of trials with the other. All subjects then received conditioning with A as the CS followed by a test to assess the extent of generalization to B. The test results are presented in Figure 10.

The figure shows that generalization was less substantial in Group A/B-I than in Group W, another instance of the perceptual learning effect demonstrated previously. But no such effect was evident in Group A/B-B; these subjects showed an aversion to Flavor B almost as profound as that displayed by the nonpreexposed subjects of Group W. Preexposure to a pair of stimuli reduces generalization between them only when they are presented in an intermixed fashion during preexposure.

Consideration of the role played by inhibitory links between the unique features of A and B supplies a ready explanation for this pattern of results. In Group W, we assume, the formation of an  $a-c$  association during conditioning allows B to evoke a CR by way of the chain  $b-c-a$ . Inhibitory links between  $b$  and  $a$ , formed during preexposure, will eliminate this source of generalized responding; such links are likely to form only with the intermixed procedure. An inhibitory link will form on a  $bc$  trial, for



**Figure 10.** Group mean scores on the test session with Flavor B after conditioning to A. Group A/B-I had received preexposure in which trials with A and B were intermixed; Group A/B-B had received separate blocks of trials with A and with B; Group W had received only water in preexposure. (Data from Symonds & Hall, 1995, Experiment 2.)

instance, only when the excitatory  $c-a$  link already has some strength; similarly,  $ac$  trials will be effective in producing inhibition only when the  $c-b$  link already exists. The alternation of A and B trials in the intermixed condition will be the optimal arrangement for ensuring that the appropriate connection has strength on each trial. With the blocked procedure, on the other hand, there is only a single transition from one trial type to the other, the excitatory connection established during the first block will extinguish during the second, and the opportunity for inhibitory links to form will be minimal. The absence of the perceptual learning effect after blocked exposure is thus just what would be expected on the basis of the account proposed by McLaren et al. (1989).

### Conclusions

These studies show that latent inhibition and inhibitory conditioning could play a role in producing the perceptual learning effect. More generally, they allow the conclusion that, although the consequences of the image conditioning process might tend to obscure the perceptual learning effect, the existence of the effect is not incompatible with the existence of the process. In the procedures used to demonstrate perceptual learning, it is possible to identify a range of other factors that will be operating, factors that will oppose the effects of image conditioning and could therefore result in a net enhancement of discrimination. Furthermore, the factors that have been identified above do not require any departure from the standard principles of associative learning and are (for the most part; see below) compatible with those assumed to be operating in image conditioning.

This conclusion requires two qualifications. First, it must be acknowledged that the explanation of perceptual learning in terms of associative learning processes is, at present, just a hypothesis and that an alternative account is available. E. J. Gibson (1969) attributes perceptual learning to a *differentiation* process that renders subjects responsive to the unique and distinguishing aspects of stimuli. The learning process involved in differentiation is said to be nonassociative in nature. This proposal does more than simply accommodate the basic facts of perceptual learning; it can, for instance, deal with the results of Symonds and Hall's (1995) Experiment 2. Differentiation is said to operate best when there is an opportunity to compare the stimuli. The mechanism by which stimulus comparison might operate is not made clear, but it seems safe to assume that comparison will be more likely with the intermixed schedule rather than the blocked schedule of stimulus exposure. To this extent, the observed results are to be expected on the basis of Gibson's theory.

Second, an important element of the associative interpretation is the proposal that an inhibitory link will be formed when the occurrence of one stimulus (e.g., A) accompanies the associative activation of the representation of another (in this case, the unique features of B). Although Wagner's (1981) theory asserts that an inhibitory link forms in just these circumstances, it is not clear (as we saw in our discussion of second-order condition-

ing) that Konorski's (1967) theory can confidently predict the development of inhibition here. If we accept Konorski's analysis, therefore, it follows that the perceptual learning effect cannot be explained in terms of inhibitory links formed between unique features of the stimuli, and some other explanation (such as that offered by E. J. Gibson, 1969) must be developed for the results reported by Symonds and Hall (1995). The alternative is to decline to accept Konorski's account in its entirety, with its assertion that an image of a stimulus can function both as a CS and as a US in association formation. This possibility is taken up in what follows.

### THEORETICAL IMPLICATIONS

We began by asking the question whether the associatively activated representation of an event (the image of that event) can serve as a substitute for the event itself in associative learning. Two views were contrasted: Konorski's (1967) theory, which makes no distinction between image and direct perception in the matter of association formation, and Wagner's (e.g., 1981) theory, which gives the image (the A2 state of activation) rather different properties from the A1 state that corresponds to direct perception of a current stimulus. In this latter theory, the only associative function of the A2 state is that it supports the acquisition of inhibitory powers by a node that is concurrently in the A1 state. Our survey of the evidence has tended to favor some version of Konorski's view. There is direct evidence that an image can come to act as a CS when it is paired with the presentation of a US, and this process helps to explain a range of phenomena demonstrated in experiments on sensory preconditioning, acquired equivalence, and perceptual learning. Mechanisms that could play a role in the backward equivalence effect and in second-order conditioning become available, if it is also allowed that an image can function as a US (i.e., that concurrent presentation of a stimulus with the evocation of the image of another can result in the formation of an excitatory link from the former to the latter). Our concern now is to assess how these effects might best be dealt with in terms of the mechanisms proposed by formal theories of associative learning.

Given what has just been said, it might be supposed that Konorski's theory would have no trouble in accommodating the phenomenon of learning about associatively activated representations. And certainly the theory has no problem with the basic fact that such a representation can come to act as a CS. But difficulties arise as soon as inhibitory learning enters the picture. As was mentioned in the discussion of second-order conditioning, Konorski construes inhibition as involving an excitatory association between the CS and a special sort of event representation, one that was referred to (in a simplification of Konorski's own analysis) as a no-US representation. In the second-order conditioning procedure, it was argued, two excitatory associations will form: one between the CS and the (associatively activated) repre-

sentation of the US, and the other between the CS and the no-US representation. The trouble is that, given this analysis, there is no reason why excitatory conditioning should give way to conditioned inhibition over the course of training—the result that is observed.

These problems could be avoided by adopting the position that, although an image can act as a CS, it cannot act as a US. This means giving up Konorski's interpretation of second-order conditioning; however, other associative mechanisms are available that can explain the basic effect. The same holds true for backward equivalence, where associations involving an image as the US constitute just one possible source of the observed results. The rest of the phenomena discussed in this article would remain as examples of conditioning with an image as the CS. Unfortunately, this solution is quite incompatible with Konorski's theory. The essence of the theory, as it concerns us here, is that the activation induced in a stimulus node by way of an associative connection does not differ in its nature from that induced by direct application of the stimulus itself. But since the state produced by direct stimulation can function, in appropriate circumstances, as either a CS or a US, then the same must hold true for the state produced by associative means. To assert that an image can function as a CS but not as a US amounts to an abandoning of the basic assumption of the theory; it requires us to accept that the direct and associative routes produce states of activation that have different properties.

This conclusion prompts a reconsideration of Wagner's (1981) account with its differentiation between A1 and A2 states. Wagner's rules for association formation assert that excitatory connections form between nodes that are both in A1 and that an inhibitory link will form between a node in A1 and one that is in A2. This latter assumption allows inhibitory learning to go on during acquisition and provides a mechanism whereby an asymptote will eventually be reached; it also ensures that extinction will occur if presentations of the US are discontinued. As it stands, this theory has no way of dealing with the phenomena described in this article, but a simple modification will allow it to do so (see Holland, 1983). The theory in its original form holds that the only connection formed between nodes in A2 and A1 is the inhibitory link from the latter to the former. But there is no compelling reason to prevent our adding the assumption that this state of affairs will also establish an excitatory link from the former to the latter. In this way, we are able to retain the various explanatory advantages that Wagner's basic theory confers and also add a mechanism whereby the image of a stimulus can come to act as excitatory CS, thus widening its explanatory power yet further. Certainly, with this amendment, the theory can accommodate the majority of the phenomena discussed in this review. General acceptance of the modified theory will, no doubt, depend on the outcome of further research. The key assumption of this interpretation is that an image can come to act as a CS but will not serve as a

US. Direct examination of the latter possibility is currently under way, and should it yield positive results, it will make necessary a reconsideration not only of Wagner's theory but of theories of association formation more generally.

## REFERENCES

- BOLHUIS, J. J., & HONEY, R. C. (1994). Within-event learning during filial imprinting. *Journal of Experimental Psychology: Animal Behavior Processes*, **20**, 240-248.
- BONARDI, C., & HALL, G. (1994). Occasion setting renders stimuli more similar: Acquired equivalence between the targets of feature-positive discriminations. *Quarterly Journal of Experimental Psychology*, **47B**, 63-81.
- BONARDI, C., REY, V., RICHMOND, M., & HALL, G. (1993). Acquired equivalence of cues in pigeon autoshaping: Effects of training with common consequences and with common antecedents. *Animal Learning & Behavior*, **21**, 369-376.
- BROGDEN, W. J. (1939). Sensory pre-conditioning. *Journal of Experimental Psychology*, **25**, 323-332.
- BROWN, C. P., & KING, M. G. (1969). Backward sensory preconditioning: The ineffectiveness of the procedure under optimal conditions. *Australian Journal of Psychology*, **21**, 55-58.
- CHAMIZO, V. D., & MACKINTOSH, N. J. (1989). Latent learning and latent inhibition in maze discriminations. *Quarterly Journal of Experimental Psychology*, **41B**, 21-31.
- CHANNELL, S., & HALL, G. (1983). Contextual effects in latent inhibition with an appetitive conditioning procedure. *Animal Learning & Behavior*, **11**, 67-74.
- COPOCK, W. J. (1958). Pre-extinction in sensory preconditioning. *Journal of Experimental Psychology*, **55**, 213-219.
- D'AMATO, M. R., SALMON, D. P., & COLOMBO, M. (1985). Extent and limits of the matching concept in monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes*, **11**, 35-51.
- FUDIM, O. K. (1978). Sensory preconditioning of flavors with a formalin-induced sodium need. *Journal of Experimental Psychology: Animal Behavior Processes*, **4**, 276-285.
- GAGNÉ, R. M., & BAKER, K. E. (1950). Stimulus predifferentiation as a factor in transfer of training. *Journal of Experimental Psychology*, **40**, 439-451.
- GARCIA, J., BRETT, L. P., & RUSINIAK, K. W. (1989). Limits of Darwinian conditioning. In S. B. Klein & R. R. Mowrer (Eds.), *Contemporary learning theories: Instrumental conditioning theory and the impact of biological constraints on learning* (pp. 181-203). Hillsdale, NJ: Erlbaum.
- GARCIA, J., & KOELLING, R. A. (1966). Relation of cue to consequence in avoidance learning. *Psychonomic Science*, **4**, 123-124.
- GIBSON, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.
- GIBSON, E. J., & WALK, R. D. (1956). The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *Journal of Comparative & Physiological Psychology*, **49**, 239-242.
- GIBSON, J. J., & GIBSON, E. J. (1955). Perceptual learning—Differentiation or enrichment? *Psychological Review*, **62**, 32-41.
- HALL, G. (1979). Exposure learning in young and adult laboratory rats. *Animal Behaviour*, **27**, 586-591.
- HALL, G. (1980). Exposure learning in animals. *Psychological Bulletin*, **88**, 535-550.
- HALL, G. (1991). *Perceptual and associative learning*. Oxford: Oxford University Press, Clarendon Press.
- HALL, G. (1994). Pavlovian conditioning: Laws of association. In N. J. Mackintosh (Ed.), *Handbook of perception and cognition: Vol. 9. Animal learning and cognition* (pp. 15-43). Orlando, FL: Academic Press.
- HALL, G., RAY, E., & BONARDI, C. (1993). Acquired equivalence between cues trained with a common antecedent. *Journal of Experimental Psychology: Animal Behavior Processes*, **19**, 391-399.
- HAYES, S. C. (1989). Nonhumans have not yet shown stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, **51**, 385-392.
- HOGAN, D. E., & ZENTALL, T. R. (1977). Backward associations in the pigeon. *American Journal of Psychology*, **90**, 3-15.
- HOLLAND, P. C. (1981). Acquisition of representation-mediated conditioned food aversions. *Learning & Motivation*, **12**, 1-18.
- HOLLAND, P. C. (1983). Representation-mediated overshadowing and potentiation of conditioned aversions. *Journal of Experimental Psychology: Animal Behavior Processes*, **9**, 1-13.
- HOLLAND, P. C. (1985). Element pretraining influences the content of appetitive serial compound conditioning in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, **11**, 367-387.
- HOLLAND, P. C. (1989). Acquisition and transfer of conditional discrimination performance. *Journal of Experimental Psychology: Animal Behavior Processes*, **15**, 154-165.
- HOLLAND, P. C. (1990). Event representation in Pavlovian conditioning: Image and action. *Cognition*, **37**, 105-131.
- HOLLAND, P. C., & FORBES, D. T. (1982a). Control of conditional discrimination performance by CS-evoked event representations. *Animal Learning & Behavior*, **10**, 249-256.
- HOLLAND, P. C., & FORBES, D. T. (1982b). Representation-mediated extinction of conditioned flavor aversions. *Learning & Motivation*, **13**, 454-471.
- HOLLAND, P. C., & RESCORLA, R. A. (1975). The effects of two ways of devaluing the unconditioned stimulus after first- and second-order appetitive conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, **1**, 355-363.
- HONEY, R. C. (1990). Stimulus generalization as a function of stimulus novelty and familiarity in rats. *Journal of Experimental Psychology: Animal Behavior Processes*, **16**, 178-184.
- HONEY, R. C., & HALL, G. (1989). The acquired equivalence and distinctiveness of cues. *Journal of Experimental Psychology: Animal Behavior Processes*, **15**, 338-346.
- HONEY, R. C., & HALL, G. (1991). Acquired equivalence and distinctiveness of cues using a sensory-preconditioning procedure. *Quarterly Journal of Experimental Psychology*, **43B**, 121-135.
- HONEY, R. C., HORN, G., & BATESON, P. (1993). Perceptual learning during filial imprinting: Evidence from transfer of training studies. *Quarterly Journal of Experimental Psychology*, **46B**, 253-269.
- HULL, C. L. (1939). The problem of stimulus equivalence in behavior theory. *Psychological Review*, **46**, 9-30.
- JACKSON, D. E., & DELPRATO, D. J. (1974). Aversive CSs suppress lever pressing for food but not the eating of free food. *Learning & Motivation*, **5**, 448-458.
- JAMES, W. (1890). *The principles of psychology*. New York: Holt.
- KONORSKI, J. (1967). *Integrative activity of the brain*. Chicago: University of Chicago Press.
- KUNO, H., KITADATE, T., & IWAMOTO, T. (1994). Formation of transitivity in conditional matching to sample by pigeons. *Journal of the Experimental Analysis of Behavior*, **62**, 399-408.
- LAWRENCE, D. H. (1949). Acquired distinctiveness of cues. I. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology*, **39**, 770-784.
- LEYLAND, C. M. (1977). Higher-order autoshaping. *Quarterly Journal of Experimental Psychology*, **29**, 607-619.
- MACKINTOSH, N. J., KAYE, H., & BENNETT, C. H. (1991). Perceptual learning in flavour-aversion conditioning. *Quarterly Journal of Experimental Psychology*, **43B**, 297-322.
- MAZUR, J. E., & WAGNER, A. R. (1982). An episodic model of associative learning. In M. L. Commons, R. J. Herrnstein, & A. R. Wagner (Eds.), *Quantitative analyses of behavior* (Vol. 3, pp. 3-39). Cambridge, MA: Ballinger.
- 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.
- MILL, J. S. (1974). *System of logic*. Toronto: University of Toronto Press. (Original work published 1843)
- MILLER, N. E. (1948). Theory and experiment relating psychoanalytic displacements to stimulus-response generalization. *Journal of Abnormal & Social Psychology*, **43**, 155-178.

- MILLER, N. E., & DOLLARD, J. (1941). *Social learning and imitation*. New Haven, CT: Yale University Press.
- NAIRNE, J. S., & RESCORLA, R. A. (1981). Second-order conditioning with diffuse auditory reinforcers in the pigeon. *Learning & Motivation*, **12**, 65-91.
- PEARCE, J. M., & HALL, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review*, **87**, 532-552.
- PREWITT, E. P. (1967). Number of preconditioning trials in sensory preconditioning using CER training. *Journal of Comparative & Physiological Psychology*, **64**, 360-362.
- RESCORLA, R. A. (1973). Second-order conditioning: Implications for theories of learning. In F. J. McGuigan & D. B. Lumsden (Eds.), *Contemporary approaches to conditioning and learning* (pp. 127-150). Washington, DC: V. H. Winston.
- RESCORLA, R. A. (1980). Simultaneous and successive associations in sensory preconditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, **6**, 207-216.
- RESCORLA, R. A. (1981). Simultaneous associations. In P. Harzem & M. Zeiler (Eds.), *Advances in analysis of behavior: Vol. 2. Predictability, correlation and contiguity* (pp. 47-79). Chichester, U.K.: Wiley.
- RESCORLA, R. A., & CUNNINGHAM, C. L. (1978). Within-compound flavor associations. *Journal of Experimental Psychology: Animal Behavior Processes*, **4**, 267-275.
- RESCORLA, R. A., & DURLACH, P. J. (1981). Within-event learning in Pavlovian conditioning. In N. E. Spear & R. R. Miller (Eds.), *Information processing in animals: Memory mechanisms* (pp. 81-111). Hillsdale, NJ: Erlbaum.
- RESCORLA, R. A., & FREBERG, L. (1978). The extinction of within-compound flavor associations. *Learning & Motivation*, **4**, 411-427.
- RESCORLA, R. A., & WAGNER, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. 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.
- ROITBLAT, H. L. (1987). *Introduction to comparative cognition*. New York: Freeman.
- SAUNDERS, K. J. (1989). Naming in conditional discrimination and stimulus equivalence. *Journal of the Experimental Analysis of Behavior*, **51**, 379-384.
- SIDMAN, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech & Hearing Research*, **14**, 5-13.
- SIDMAN, M. (1990). Equivalence relations: Where do they come from? In D. E. Blackman & H. Lejeune (Eds.), *Behaviour analysis in theory and practice: Contributions and controversies* (pp. 93-114). Hove, U.K.: Erlbaum.
- SIDMAN, M., & TAILBY, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, **37**, 5-22.
- SIEGEL, S. (1967). Overtraining and transfer processes. *Journal of Comparative & Physiological Psychology*, **64**, 471-477.
- SILVER, C. A., & MEYER, D. R. (1954). Temporal factors in sensory preconditioning. *Journal of Comparative & Physiological Psychology*, **47**, 57-59.
- SPETCH, M. L., WILKIE, D. M., & PINEL, J. P. J. (1981). Backward conditioning: A reevaluation of the empirical evidence. *Psychological Bulletin*, **89**, 163-175.
- STROMER, R., & OSBORNE, J. G. (1982). Control of adolescents' arbitrary matching-to-sample by positive and negative stimulus relations. *Journal of the Experimental Analysis of Behavior*, **37**, 329-348.
- SUTHERLAND, N. S., & MACKINTOSH, N. J. (1971). *Mechanisms of animal discrimination learning*. New York: Academic Press.
- SWARTZENTRUBER, D. (1995). Modulatory mechanisms in Pavlovian conditioning. *Animal Learning & Behavior*, **23**, 123-143.
- SYMONDS, M., & HALL, G. (1995). Perceptual learning in flavor aversion conditioning: Roles of stimulus comparison and latent inhibition of common stimulus elements. *Learning & Motivation*, **26**, 203-219.
- SYMONDS, M., & HALL, G. (1996). *Comparison and learning about common stimulus elements*. Manuscript submitted for publication.
- TAIT, R. W., MARQUIS, H. A., WILLIAMS, R., WEINSTEIN, L., & SUBOSKI, M. D. (1969). Extinction of sensory preconditioning. *Journal of Comparative & Physiological Psychology*, **69**, 170-172.
- TROBALON, J. B., CHAMIZO, V. D., & MACKINTOSH, N. J. (1992). Role of context in perceptual learning in maze discriminations. *Quarterly Journal of Experimental Psychology*, **44B**, 57-73.
- TROBALON, J. B., SANSO, J., CHAMIZO, V. D., & MACKINTOSH, N. J. (1991). Perceptual learning in maze discriminations. *Quarterly Journal of Experimental Psychology*, **43B**, 389-402.
- URCUIOLI, P. J., & ZENTALL, T. R. (1993). A test of comparison stimulus substitutability following one-to-many matching by pigeons. *Psychological Record*, **43**, 745-759.
- URCUIOLI, P. J., ZENTALL, T. R., & DEMARSE, T. (1995). Transfer to derived sample-comparison relations by pigeons following many-to-one versus one-to-many matching with identical training relations. *Quarterly Journal of Experimental Psychology*, **48**, 158-178.
- 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. (1989). Evolution of a structured connectionist model of Pavlovian conditioning (AESOP). In S. B. Klein & R. R. Mowrer (Eds.), *Contemporary learning theories: Pavlovian conditioning and the status of traditional learning theory* (pp. 149-189). Hillsdale, NJ: Erlbaum.
- WAGNER, A. R., & LAREW, M. B. (1985). Opponent processes and Pavlovian inhibition. In R. R. Miller & N. E. Spear (Eds.), *Information processing in animals: Conditioned inhibition* (pp. 233-265). Hillsdale, NJ: Erlbaum.
- WARD-ROBINSON, J. (1994, April). *Evidence for an associative account of acquired equivalence*. Paper presented at the meeting of the Experimental Analysis of Behaviour Group, London.
- WARD-ROBINSON, J., & HALL, G. (in press). Backward sensory preconditioning. *Journal of Experimental Psychology: Animal Behavior Processes*.
- YIN, H., BARNET, R. C., & MILLER, R. R. (1994). Second-order conditioning and Pavlovian conditioned inhibition: Operational similarities and differences. *Journal of Experimental Psychology: Animal Behavior Processes*, **20**, 419-428.
- ZENTALL, T. R., SHERBURNE, L. M., STEIRN, J. N., RANDALL, C. K., ROPER, K. L., & URCUIOLI, P. J. (1992). Common coding in pigeons: Partial versus total reversals of one-to-many conditional discriminations. *Animal Learning & Behavior*, **20**, 373-381.

## NOTES

1. A point of terminology: The term US often refers to the motivationally significant event that is preceded by presentation of the neutral CS in the standard conditioning procedure. But any effective stimulus is a US (whether or not it evokes an overt response) in that it will directly evoke activity in its own node; any event that acquires the ability to evoke activity in some other node by way of an associative link can be regarded as a CS. There is some potential for ambiguity here, but I trust this will always be resolved, in the present article, by the context in which these terms are used.

2. Further evidence comes from the finding (e.g., Fudim, 1978) that sensory preconditioning can be obtained when there is no separate conditioning phase (the value of one of the elements of the compound being changed not by associative means but by a manipulation of the motivational state of the animal). It is difficult to explain this effect in terms of an image conditioning process.

3. This is to simplify Konorski's (1967) account. He assumes not a single US representation but separate centers encoding the sensory properties of the event and the affective state it engenders. Each of these has its appropriate negative counterpart.

4. Such overshadowing will also occur during reinforcement of A, after A and B have undergone equivalence training. But since the c elements will suffer at the expense of others (the X stimulus) that are common to A and B, an enhancement of generalization can still be predicted.

(Manuscript received July 3, 1995;

revision accepted for publication October 19, 1995.)