

# Discrimination blocking: Acquisition versus performance deficits in human contingency learning

LEYRE CASTRO AND EDWARD A. WASSERMAN  
University of Iowa, Iowa City, Iowa

We compared acquisition and performance accounts of human contingency learning. After solving a discrimination in Phase 1, in which Cue A predicted the occurrence of the outcome and Cue B predicted its nonoccurrence ( $A+/B-$ ), a new discrimination ( $X+/Y-$ ) was superimposed in Phase 2 ( $AX+/BY-$ ). The participants were finally trained in Phase 3 with the added discrimination, which either maintained the same contingencies as those in Phase 2 ( $X+/Y-$ ; Experiment 1) or reversed the contingencies ( $X-/Y+$ ; Experiment 2). According to competitive-learning theories (e.g., Rescorla & Wagner, 1972), there should be no learning of the added discrimination in Phase 2, so that no advantage or disadvantage for this discrimination should be observed in Phase 3. In contrast, performance theories, such as the comparator hypothesis (Miller & Matzel, 1988), contend that learning of the added discrimination in Phase 2 should proceed normally; so, in Phase 3, an advantage for the added discrimination should be observed in Experiment 1, but a disadvantage should be observed in Experiment 2. Our participants learned about the added discrimination and generally showed the effects predicted by the comparator hypothesis.

When a cue fails to provide new information about the contingencies between events in the environment, organisms seem to disregard it. Perhaps because processing resources are limited, the lack of concern for redundant or uninformative cues appears to contribute to the efficient management of those resources. This informational analysis is nicely illustrated by the *blocking effect*.

Kamin (1968) observed that initial training with one cue (Cue A) considerably reduced performance of a conditioned response to a new cue (Cue X) that was added to the original cue in a second phase of training. According to Kamin, the animals might notice the presence of Cue X but not condition to it because the cue was redundant. That is, in the second phase of training, the animals already expect the occurrence of the outcome on the basis of the presence of Cue A. Cue X does not signal anything different from Cue A. Therefore, "the 'mental work' necessary for the formation of an association between the CS [conditioned stimulus or cue] and the US [unconditioned stimulus or outcome] will not be provoked" (Kamin, 1968, p. 30).

Contemporary theories of associative learning have incorporated this idea (for a review, see, e.g., Allan, 1993; Shanks, 1995; Young & Wasserman, 2004). In their highly influential learning model, Rescorla and Wagner (1972; henceforth, RW) mathematically formalized the notion that learning of a cue–outcome relationship depends on the degree to which the outcome is surprising. When the outcome is already predicted by other cues, no learning will occur to the added cue. According to RW, there is a

maximal level of learning that an outcome can support. When Cue A is initially paired on its own with the outcome, the associative strength of this cue will gradually grow until it reaches its asymptote. At this point, Cue A will fully predict the outcome, and no more associative strength will be available for any other cues; such added cues will not enter into association with the outcome.

Other models entail similar principles, although they differ in how events are processed. For example, Mackintosh (1975b; see also McLaren & Mackintosh, 2000, 2002; Pearce & Hall, 1980) proposed that when the outcome is not surprising, an added cue does not provide any new information, so that organisms will pay no more attention to it. Because organisms learn to ignore the added cue, the cue–outcome relationship will not be learned.

Despite their technical disparities, all of these models assume that an added cue will not acquire associative strength when the outcome is already fully predicted by other cues. These models focus on the processes that occur in training and contend that learning is a competitive affair in which only a subset of available cues will become associated with the outcome; just those cues that predict something novel about the contingencies of reinforcement will prevail. Thus, we will refer to these models as *competitive-learning* models.

In contrast, other models do not consider that organisms' learning capacities are so constrained or that learning requires novelty and surprise. Instead, the critical processes that control behavior occur when, after learning,

a response must be performed. For example, Miller and his colleagues (see Denniston, Savastano, & Miller, 2001; Miller & Matzel, 1988) hypothesized that organisms learn a cue–outcome association as long as there is contiguity between the events; the surprisingness of the outcome does not play a critical role in learning. The absence of a response to a cue—as in the blocking effect—need not be due to the lack of an association between the cue and the outcome, but to the failure to express that cue–outcome association. We will refer to these models as *performance* models.

According to the comparator hypothesis (henceforth CH; Denniston et al., 2001; Miller & Matzel, 1988), a conditioned response to the target cue results from a comparison process between the direct activation of the outcome by the target cue and the indirect activation of the outcome by other cues—so-called *comparators*—that had earlier been paired with the target cue. If the indirect activation by the comparators is stronger than the direct activation by the target cue, a weak response to the target cue will be observed. This weak response does not mean that an association between the target cue and the outcome does not exist; it merely means that there is a stronger association of another cue with the outcome that successfully competes for the activation of the outcome and, consequently, for the evocation of the response.

In the last decade, several efforts have been made to ascertain whether a learning or a performance deficit is responsible for blocking. Williams (1996) first trained rats with the standard two-phase blocking design: Cue A followed by the outcome in Phase 1, and Cue X in compound with Cue A followed by the outcome in Phase 2. Williams reasoned that if the added Cue X is effectively associated with the outcome, then when Cue X is later presented along with novel Cue Y, Cue X should be able to block the response to Cue Y in subsequent testing. However, Cue Y did elicit a robust conditioned response in Williams's experiments. Cue X did not appear to be able to produce a decrement in responding to Cue Y, presumably because there had been no prior learning of the Cue X–outcome association. A learning deficit here seemed to be responsible for the blocking effect (also see Rauhut, McPhee, & Ayres, 1999).

Blaisdell, Gunther, and Miller (1999) adopted a different approach. They argued that learning of the Cue X–outcome association had occurred during training in compound with Cue A but that this learning did not translate into performance because a stronger association of Cue A with the outcome prevented the expression of the Cue X–outcome association. So Blaisdell et al. extinguished the Cue A–outcome association after the compound-training phase and before the testing of Cue X. A strong conditioned response to Cue X was then observed. Blaisdell et al. concluded that their rats had learned the added Cue X–outcome association during the compound-training phase. Here, the blocking effect seemed to be due to a performance deficit.

To further complicate the matter, although the results in Blaisdell et al. (1999) were inconsistent with the original RW model, they still could be understood as the result of

a competitive-learning process if it was assumed that the associative strengths of *absent* cues change in the direction *opposite* to those of present cues (Dickinson & Burke, 1996; Van Hamme & Wasserman, 1994). Furthermore, Denniston, Savastano, Blaisdell, and Miller (2003) found that after pairing Compound AX with the outcome, Cue X was overshadowed and later failed to block acquisition of a novel added cue, Cue Y, when both were paired with the outcome (XY–outcome). This result is equivalent to Williams's (1996) findings. The overshadowed Cue X could not block acquisition of Cue Y, so that a learning deficit seemed to be involved. But Denniston et al. (2003) did something else. They extinguished the overshadowing Cue A after XY training and observed that, after extinction of Cue A, responding to Cue Y was impaired. Therefore, Cue X could *retrospectively* block responding to Cue Y, provided that the comparator for Cue X (i.e., Cue A) no longer had a strong association with the outcome. This finding cannot be explained by the revised acquisition models (Dickinson & Burke, 1996; Van Hamme & Wasserman, 1994) or by the original CH. However, it can be accommodated by an extension of the original CH, which holds that responding to a target cue is influenced not only by its direct comparators, but also by higher order comparators—cues that have not been directly paired with the target cues, but with the comparator cues (see Denniston et al., 2003; Denniston et al., 2001).

In the present study, we further explored whether human participants' contingency-learning performance more clearly reflects the involvement of acquisition or performance mechanisms. In order to do so, we used a different experimental tactic. Imagine a situation in which participants are trained to solve a *discrimination* in which Cue A predicts the occurrence of the outcome, whereas Cue B predicts the nonoccurrence of the outcome. Once this discrimination is learned, a new discrimination is added, so that Cue A and Cue X in compound predict the presence of the outcome, whereas Cue B and Cue Y in compound predict the absence of the outcome. Will participants learn about the added X+/Y– discrimination?

Johnson and Cumming (1968) presented pigeons with this kind of task. The birds had to peck a green key with a vertical line, which was followed by food, and to withhold pecking a red key with a horizontal line, which was never followed by food. When single elements of the discrimination were trained before the compound discrimination (e.g., the green key alone reinforced and the red key alone nonreinforced before the green key plus the vertical line reinforced and red key plus the horizontal line nonreinforced), it was difficult for the pigeons to learn the superimposed discrimination (vertical line vs. horizontal line). Johnson and Cumming suggested that this difficulty was due to a lack of attention to the added elements. Nonetheless, these results can also be interpreted in the same way as the forward-blocking effect. According to RW, once the birds had learned the single discrimination to an asymptotic level—so that the green key perfectly predicted the occurrence of food and the red key perfectly predicted the nonoccurrence of food—there should be no more associative strength available for the elements of the added

discrimination to become excitatory or inhibitory, so no learning should take place. In contrast, the CH would say that the added discrimination should be learned but, because the other single discrimination is stronger, the added discrimination should fail to control performance.

Taking Johnson and Cumming’s (1968) design as a starting point, we explored whether training human participants with an element discrimination (e.g., A+/B−) before training them with a compound discrimination (e.g., AX+/BY−) would impair the learning or simply the expression of the added element discrimination (e.g., X+/Y−).

**EXPERIMENT 1**

We first trained our participants to solve a discrimination between Cue A, always followed by the outcome, and Cue B, never followed by the outcome (i.e., A+/B−). Later, two new cues, X and Y, were added to the initial discriminative cues, A and B. In the *consistent* condition, Cue X was always paired with A+, whereas Cue Y was always paired with B−. In the *inconsistent* condition, both X and Y were equally often paired with A+ and B− (the experimental design is summarized in Table 1). Thus, a solvable discrimination, X+/Y−, was added in the consistent condition, whereas an unsolvable discrimination, X±/Y±, was added in the inconsistent condition.

According to RW, initial A+ and B− training to asymptote should completely block learning about any added cues—here, X and Y. So, although a solvable discrimination, X+/Y−, was presented in the consistent condition and an unsolvable discrimination, X±/Y±, was presented in the inconsistent condition, in neither of the conditions should the participants actually learn about Cues X and Y. Hence, after Phase 2 training, RW would predict no associative differences between Cues X and Y in the consistent or the inconsistent condition; accordingly, no differential discriminative performance should be observed in either of the conditions.

On the contrary, according to CH, the participants should learn about all of the cues that were presented at any given time via spatial and temporal contiguity. This learning might not always be expressed, but it should take place nevertheless. Hence, the participants should be able to learn an X+/Y− discrimination in the consistent condition, but not in the inconsistent condition. Of course, test-

ing Cues X and Y at this time might yield similar responses in both conditions, due to the presence of two highly effective comparators, Cues A and B, which might not allow the expression of a learned discrimination between Cues X and Y. So, at this point, it might not be possible to distinguish between learning and performance models; both RW and CH would predict similar responding to Cues X and Y in the consistent and the inconsistent conditions.

To determine whether or not the participants had, in fact, learned a discrimination between Cues X and Y in Phase 2, we included a third phase in which these cues were separately trained: X alone, followed by the presence of the outcome, and Y alone, followed by the absence of the outcome. If the participants had learned about Cues X and Y in Phase 2, Phase 3 learning should proceed more quickly in the consistent condition, which entailed a solvable Phase 2 discrimination, than in the inconsistent condition, which entailed an unsolvable Phase 2 discrimination. On the contrary, if the participants had not learned about Cues X and Y in Phase 2, Phase 3 learning should proceed similarly in both conditions.

**Method**

**Participants, Apparatus, and Stimuli.** A total of 43 undergraduate students at the University of Iowa received course credit for their voluntary participation. The experiment was conducted on four iMac computers. From 1 to 4 participants were studied concurrently on identically configured computer workstations. Instructions and task stimuli were presented on the screen, and the participants were required to make their responses by using the mouse or the computer keyboard.

Pictures of 18 different foods (cherries, mushrooms, strawberries, fish, apples, meat, avocados, bananas, ice cream, nuts, peppers, broccoli, lemons, grapes, toast, tomatoes, carrots, and coffee) served as cues. Those cues were counterbalanced following a between-subjects Latin square design, which ensured that each food was equally often assigned to each cue role. The outcomes were *allergic reaction* on reinforced trials and *no allergic reaction* on nonreinforced trials.

**Procedure.** The participants were introduced to a scenario in which they played the role of an allergist trying to discover which foods would or would not cause an allergic reaction in a specific individual, Mr. X (see the Appendix for the instructions). To perform the task, the participants had to study different daily allergy tests, in which Mr. X had or had not suffered an allergic reaction after eating certain foods. On each learning trial, pictures of one or two foods appeared centered at the top of the screen. When two foods were presented, each specific picture appeared on half of the trials to the left

**Table 1**  
**Design of Experiments 1 and 2**

Condition	Phase 1	Phase 2	Phase 3	
			Experiment 1	Experiment 2
Consistent	16 A <sub>1</sub> +, 16 B <sub>1</sub> −	8 A <sub>1</sub> X <sub>1</sub> +, 8 B <sub>1</sub> Y <sub>1</sub> −	8 X <sub>1</sub> +, 8 Y <sub>1</sub> −	8 X <sub>1</sub> −, 8 Y <sub>1</sub> +
		8 E <sub>1</sub> F <sub>1</sub> +, 8 G <sub>1</sub> H <sub>1</sub> −	8 F <sub>1</sub> +, 8 H <sub>1</sub> −	8 F <sub>1</sub> −, 8 H <sub>1</sub> +
Inconsistent	16 A <sub>2</sub> +, 16 B <sub>2</sub> −	4 A <sub>2</sub> X <sub>2</sub> +, 4 B <sub>2</sub> Y <sub>2</sub> −	8 X <sub>2</sub> +, 8 Y <sub>2</sub> −	8 X <sub>2</sub> −, 8 Y <sub>2</sub> +
		4 A <sub>2</sub> Y <sub>2</sub> +, 4 B <sub>2</sub> X <sub>2</sub> −		
		4 E <sub>2</sub> F <sub>2</sub> +, 4 G <sub>2</sub> H <sub>2</sub> −	8 F <sub>2</sub> +, 8 H <sub>2</sub> −	8 F <sub>2</sub> −, 8 H <sub>2</sub> +
		4 E <sub>2</sub> H <sub>2</sub> +, 4 G <sub>2</sub> F <sub>2</sub> −	8 W <sub>2</sub> +, 8 Z <sub>2</sub> −	8 W <sub>2</sub> −, 8 Z <sub>2</sub> −

Note—The different letters represent the different cues. The + denotes that the outcome occurs, and the − denotes that the outcome does not occur.

and on half of the trials to the right. After 3 sec, the participants were required to predict whether Mr. X would develop an allergic reaction. Once the participants clicked *yes* or *no*, the actual outcome appeared below. After 3 sec, the participants could proceed to the next trial.

All of our experimental and control conditions were presented in a within-subjects manner. In the first training phase, the participants observed four elemental cues that were presented 16 times each. Cues  $A_1$  and  $A_2$  were always paired with the outcome, whereas Cues  $B_1$  and  $B_2$  were never paired with the outcome (i.e.,  $A_1+/B_1-$  and  $A_2+/B_2-$ ). The 48 trials in Phase 1 were presented in a block-randomized order.

In the second training phase, Cues  $X_1$  and  $Y_1$  were added to Cues  $A_1$  and  $B_1$ , respectively, so that the participants were presented with Compound  $A_1X_1$  always paired with the outcome and Compound  $B_1Y_1$  never paired with the outcome (i.e.,  $A_1X_1+/B_1Y_1-$ ). Therefore, a perfect contingency was presented between Cue  $X_1$  and the occurrence of the outcome and between Cue  $Y_1$  and the nonoccurrence of the outcome. Each of these two types of trials was presented eight times in Phase 2. As well, Cues  $X_2$  and  $Y_2$  were added to Cues  $A_2$  and  $B_2$ . In this case, however, pairings of Cues  $X_2$  and  $Y_2$  with outcome occurrence and nonoccurrence were inconsistent; half of the trials entailed  $X_2$  and  $Y_2$  paired with the presence of the allergic reaction, whereas half of the trials entailed  $X_2$  and  $Y_2$  paired with the absence of the allergic reaction, yielding four types of trials:  $A_2X_2+$ ,  $A_2Y_2+$ ,  $B_2X_2-$ , and  $B_2Y_2-$ . Each of these four types of trials was presented four times in Phase 2.

We also included a control discrimination, in which the training stimuli occurred in compound and none of them had previously been seen alone: Compound  $E_1F_1$  was always paired with the occurrence of outcome, whereas Compound  $G_1H_1$  was always paired with the nonoccurrence of the outcome (i.e.,  $E_1F_1+$  and  $G_1H_1-$ ). Each of these two types of trials was presented eight times. Finally, we also included an inconsistent version of novel compound training:  $E_2F_2+$ ,  $E_2H_2+$ ,  $G_2F_2-$ , and  $G_2H_2-$ , with these four types of trials presented four times each. In total, there were 64 trials in Phase 2, presented in a block-randomized order.

In the third phase of training, the elemental cues involved in the target and control discriminations were presented on their own, one associated with the outcome and the other not:  $X_1+/Y_1-$ ,  $X_2+/Y_2-$ ,  $F_1+/H_1-$ , and  $F_2+/H_2-$ . We included a novel discrimination as well,  $W+/Z-$ , to provide a baseline for learning. These 10 types of trials were presented eight times each. The 80 trials of Phase 3 were presented in a block-randomized order.

If learning of the solvable discrimination added in Phase 2,  $X_1+/Y_1-$ , was blocked by the earlier Phase 1 learning of  $A_1+/B_1-$ , the speed of learning the  $X_1+/Y_1-$  discrimination in Phase 3 should be similar to the speed of learning the  $X_2+/Y_2-$  discrimination; as well, there should be no advantage for the consistent condition over the inconsistent condition. But if the participants did learn about  $X_1+/Y_1-$  (and about  $X_2\pm/Y_2\pm$ ) in Phase 2, it should be easier for them to master the  $X_1+/Y_1-$  discrimination than the  $X_2+/Y_2-$  discrimination in Phase 3; that is, there should be an advantage for the consistent condition over the inconsistent condition.

We also compared learning of the  $X_1+/Y_1-$  discrimination with learning of the  $F_1+/H_1-$  discrimination in Phase 3. In Phase 2, the  $F_1+/H_1-$  discrimination was presented in compound with the  $E_1+/G_1-$  discrimination, but each discrimination was equally novel at that time. In this case, both CH and RW would expect learning of the  $F_1+/H_1-$  and  $E_1+/G_1-$  discriminations; these discriminations might not be as strong as if they had been presented elementally, but there should be substantial learning nonetheless. So, at the beginning of Phase 3, the  $F_1+/H_1-$  discrimination should start with some advantage over a blocked or a novel discrimination. Hence, the  $F_1+/H_1-$  discrimination should be learned more quickly than the  $X_1+/Y_1-$  discrimination if learning of the  $X_1+/Y_1-$  discrimination was blocked. On the contrary, if the participants did learn about the added  $X_1+/Y_1-$  discrimination in Phase 2, learning of the  $X_1+/Y_1-$  and the  $F_1+/H_1-$  discriminations should proceed similarly. As well, learning of both the  $X_1+/Y_1-$  and the  $F_1+/H_1-$

discriminations should proceed more quickly than learning of the novel discrimination,  $W+/Z-$ .

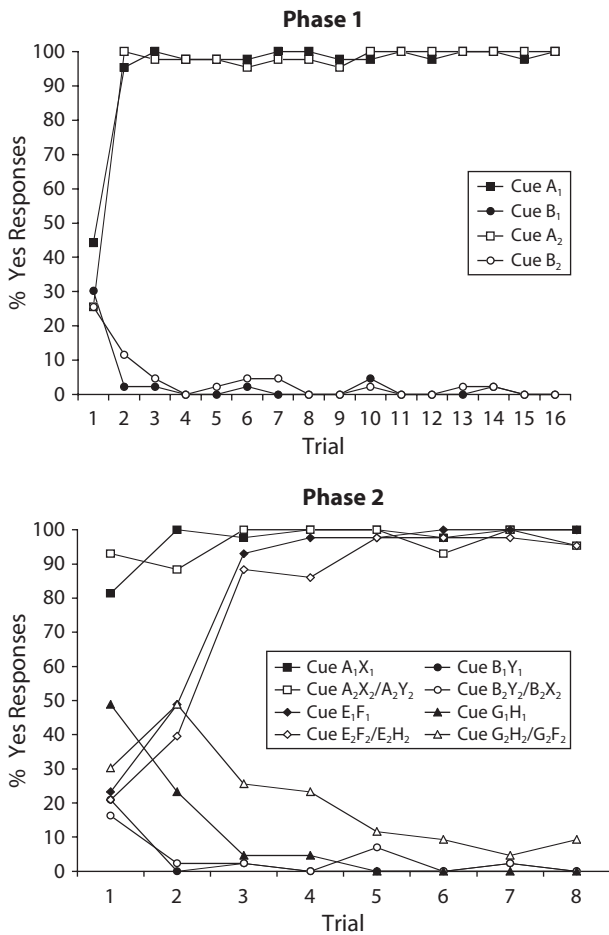
After Phase 3, the participants had to judge the extent to which each of the cues affected the likelihood of an allergic reaction in Mr. X. The 18 cues were presented in a random order. Each cue was presented on a separate screen, along with a rating scale ranging from  $-100$  (*makes very unlikely*) to  $+100$  (*makes very likely*). The middle point of the scale, 0, was labeled *does not affect likelihood*. The participants had to give their ratings by moving a slider along this scale. The initial position of the slider was set at 0 for each of the cues. Once the participants had completed the rating phase, they were debriefed and allowed to depart.

**Statistical analyses.** All of the inferential statistical analyses focused on the trial-by-trial prediction scores that the participants made in Phase 3 plus the one-time cue-rating scores that they made after Phase 3; these two sets of scores most effectively assayed any potential blocking and overshadowing effects in Phase 2. Statistical significance was set at  $\alpha = .05$ . When needed, Bonferroni correction was used to set the alpha level.

## Results and Discussion

**Training: Phases 1 and 2.** Acquisition behavior in the first and second phases of training is illustrated in Figure 1. The participants very rapidly and virtually perfectly learned the  $A_1+/B_1-$  and  $A_2+/B_2-$  discriminations that were given in Phase 1 (Figure 1, top). Beyond the second trial with each of the four cues, the participants exhibited asymptotic discriminative performance; the participants correctly predicted the occurrence of the outcome after Cues  $A_1$  and  $A_2$ , as well as the nonoccurrence of the outcome after Cues  $B_1$  and  $B_2$ . This high level of discriminative performance to each of the four cues was maintained throughout the final 14 trials of Phase 1. According to RW, if the participants learned to perfectly predict the occurrence and nonoccurrence of the outcome by the end of Phase 1, no untapped associative strength should be available for them to learn about any added cues in Phase 2. Therefore, the participants should learn nothing at all about the  $X_1+/Y_1-$  and  $X_2\pm/Y_2\pm$  contingencies that were added in Phase 2.

We should note that our assertion that the participants in our experiment had reached the asymptote by the end of Phase 1 training is based on their trial-by-trial *yes/no* predictions. It might be argued that, even when all of the participants were correctly predicting the presence and absence of the outcome at the end of Phase 1, asymptotic associative strength might not have been reached. To further explore whether or not associative strength was asymptotic after Phase 1 training, we examined our other performance measure—contingency ratings—in a different collection of research participants. We did not include contingency ratings after Phase 1 in Experiment 1, because making such ratings might have influenced later discrimination learning. So we presented two new groups of participants with the same training as that in Phase 1 (i.e.,  $A_1+/B_1-$  and  $A_2+/B_2-$ ): One group was presented with 16 trials of each type (Group 16), and the other group was presented with 32 trials of each type (Group 32). In both groups, all of the participants perfectly predicted the presence and absence of the outcome at the end of training, just as in Experiment 1 (and Experiment 2). In addition, at the end of the investigation, the participants



**Figure 1.** Percentages of predictions of the occurrence of the outcome after each of the cues in Phase 1 (top) and Phase 2 (bottom) of Experiment 1. Error bars indicate the standard error of the mean.

had to rate the extent to which each of the cues affected the likelihood of the outcome. Mean ratings for the cues paired with the presence and absence of the outcome were 96.7 and  $-82.7$ , respectively, in Group 16 and 95.2 and  $-81.2$ , respectively, in Group 32. Critically, training that was twice as long as that in our reported study did *not* lead to more extreme ratings. Hence, there is no reason to believe, in light of this evidence, that asymptotic associative strength had not been reached.

In Phase 2 (see Figure 1, bottom), although there was a small decrement in discriminative performance on the first trial with each compound cue, the participants' accuracy was very high in the consistent condition; they correctly predicted the occurrence of the outcome after  $A_1X_1$  and the nonoccurrence of the outcome after  $B_1Y_1$ . In the inconsistent condition, the participants also accurately predicted the occurrence of the outcome after  $A_2X_2$  and  $A_2Y_2$  trials, as well as the nonoccurrence of the outcome after  $B_2X_2$  and  $B_2Y_2$  trials; the addition of the unsolvable  $X_2-Y_2$  discrimination did not adversely affect the participants' performance of the  $A_2+/B_2-$  discrimination. There

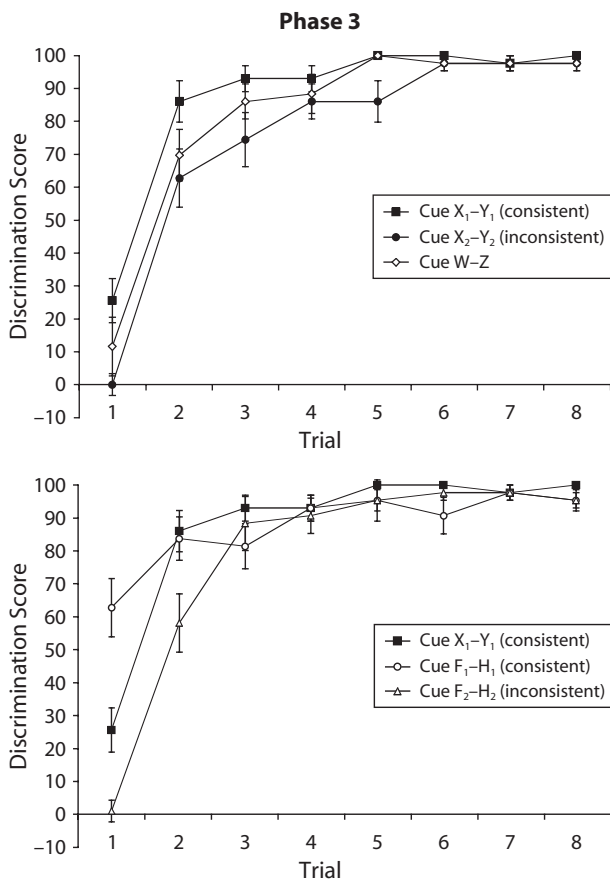
were no clear differences in discriminative performance between the consistent and the inconsistent conditions.

The compound control discriminations,  $E_1F_1+/G_1H_1-$  and  $E_2F_2+, E_2H_2+/G_2F_2-, G_2H_2-$ , generally proceeded more slowly than the discriminations containing only elements that had been presented in Phase 1 (see Figure 1, right). In addition,  $E_1F_1+/G_1H_1-$  discrimination learning was faster than  $E_2F_2+, E_2H_2+/G_2F_2-, G_2H_2-$  discrimination learning, perhaps because of the inconsistent F-outcome and H-outcome relationships. This disparity in discrimination learning between the consistent and the inconsistent conditions did not occur in the case of the  $A_1X_1+/B_1Y_1-$  and the  $A_2X_2+, A_2Y_2+/B_2X_2-, B_2Y_2-$  discriminations, perhaps because, in Phase 2, the participants relied on what they had learned about Cues  $A_1, B_1, A_2,$  and  $B_2$  in Phase 1. Nonetheless, by the fifth trial, the participants were able to accurately predict the occurrence and nonoccurrence of the outcome after each of the compound cues.

**Training: Phase 3.** To simplify the analysis of learning of the different discriminations in Phase 3, the key learning phase, we calculated discrimination scores for each of the discriminations on every trial. From the percentage of *yes* responses for each cue paired with the occurrence of the outcome, we subtracted the percentage of *yes* responses for the corresponding cue paired with the nonoccurrence of the outcome. Mean scores near 100 represent excellent discrimination (with *yes* responses close to 100 for the positive cue and with *yes* responses close to 0 for the negative cue), mean scores near 0 represent no discrimination (with *yes* responses similar for the positive and the negative cues), and mean scores near  $-100$  represent reversed discrimination (with *yes* responses close to 0 for the positive cue and with *yes* responses close to 100 for the negative cue). We obtained five different discrimination scores in Phase 3:  $X_1-Y_1, X_2-Y_2, F_1-H_1, F_2-H_2,$  and  $W-Z$ .

The development of the consistent discrimination  $X_1-Y_1$ , as compared with the inconsistent and novel discriminations, is displayed in Figure 2 (top), whereas the development of the same consistent discrimination  $X_1-Y_1$ , as compared with the control discriminations without Phase 1 training, is displayed in Figure 2 (bottom). The critical comparison concerns  $X_1-Y_1$  and  $X_2-Y_2$  (Figure 2, top). Here, there was a clear learning advantage for the consistent condition; the participants were faster to master the  $X_1-Y_1$  discrimination than the  $X_2-Y_2$  discrimination, suggesting that they had indeed learned about these cues when they were added in Phase 2, as predicted by CH. Moreover, the order of the discrimination scores,  $X_1-Y_1 > W-Z > X_2-Y_2$ , suggests an enhancement of the  $X_1-Y_1$  discrimination, as well as an impairment of the  $X_2-Y_2$  discrimination, as compared with the novel element discrimination,  $W-Z$ .

The right panel of Figure 2 shows similar learning of the  $X_1-Y_1$  and  $F_1-H_1$  discriminations. There was a disparity in discriminative performance on the first trial, suggesting that Phase 2 blocking was stronger than Phase 2 overshadowing (as predicted by both RW and CH); but beyond the second trial, prediction scores were similar for the  $X_1-Y_1$  and  $F_1-H_1$  discriminations, if not a bit



**Figure 2.** Percentage of predictions of the occurrence of the outcome in Phase 3 of Experiment 1. Discrimination scores are plotted. **Top:** The score for the consistent discrimination  $X_1-Y_1$  is compared with those for the inconsistent and novel discriminations. **Bottom:** The score for the consistent discrimination  $X_1-Y_1$  is compared with those for the control discriminations without training in Phase 1. Error bars indicate the standard error of the mean.

higher for the  $X_1-Y_1$  discrimination. Hence, learning of the *blocked* discrimination proceeded at least as well as learning of the *overshadowed* discrimination, indicative of effective learning of the  $X_1-Y_1$  discrimination in Phase 2. Finally, learning of the  $F_1-H_1$  discrimination in the consistent condition was faster than learning of the  $F_2-H_2$  discrimination in the inconsistent condition; this disparity parallels the advantage of the  $X_1-Y_1$  discrimination over the  $X_2-Y_2$  discrimination.

A 5 (discrimination:  $X_1-Y_1$ ,  $X_2-Y_2$ ,  $F_1-H_1$ ,  $F_2-H_2$ , or  $W-Z$ )  $\times$  7 (Trial: 2-8)<sup>1</sup> ANOVA of the Phase 3 training data revealed significant main effects of discrimination [ $F(4,168) = 2.48$ ,  $MS_e = 0.14$ ] and trial [ $F(6,252) = 22.13$ ,  $MS_e = 0.08$ ]. The discrimination  $\times$  trial interaction was also significant [ $F(24,1008) = 1.99$ ,  $MS_e = 0.63$ ], suggesting that learning emerged differently among the various discrimination conditions.

Follow-ups of this ANOVA concentrated on particular relevant comparisons. Linear trend analyses disclosed that the  $X_1-Y_1$  discrimination rose more quickly than the  $X_2-Y_2$  discrimination [ $F(1,42) = 8.91$ ,  $MS_e = 0.09$ ];

this outcome confirms more rapid learning in the consistent than in the inconsistent condition. If learning of the  $X_1-Y_1$  discrimination had been blocked in Phase 2, the speed of learning the  $X_1-Y_1$  discrimination in Phase 3 should have been similar to the speed of learning the  $X_2-Y_2$  discrimination. But we found an advantage for the consistent condition over the inconsistent condition. This advantage supports the idea that the participants did learn about the  $X_1$ -outcome and  $Y_1$ -no-outcome contingencies in Phase 2, as is predicted by CH, but not by RW.

Linear trend analyses also disclosed that the  $F_1-H_1$  discrimination in the consistent condition rose more quickly than the  $F_2-H_2$  discrimination in the inconsistent condition [ $F(1,42) = 3.34$ ,  $MS_e = 0.11$ ]. This difference parallels the advantage of  $X_1-Y_1$  over  $X_2-Y_2$ , again suggesting that learning in Phase 2 was similar for these discriminations ( $X_1-Y_1$  and  $F_1-H_1$ , on the one hand, vs.  $X_2-Y_2$  and  $F_2-H_2$ , on the other hand), whether or not their associates had previously been trained alone. At the same time, there was no disparity between the  $X_1-Y_1$  and the  $F_1-H_1$  discriminations [ $F(1,42) = 1.70$ ,  $MS_e = 0.10$ ] or between the  $X_2-Y_2$  and the  $F_2-H_2$  discriminations [ $F(1,42) = 0.37$ ,  $MS_e = 0.13$ ], confirming that learning proceeded similarly for each of the discriminations in the consistent condition ( $X_1-Y_1$  and  $F_1-H_1$ ) and for each of the discriminations in the inconsistent condition ( $X_2-Y_2$  and  $F_2-H_2$ ). This equivalence again suggests that similar learning occurred in Phase 2 for the  $X_1-Y_1$  discrimination, whose associates,  $A_1$  and  $B_1$ , had been individually trained in Phase 1, and for the  $F_1-H_1$  discrimination, whose associates,  $E_1$  and  $G_1$ , had not previously been trained. Likewise, similar learning occurred in Phase 2 for the  $X_2-Y_2$  discrimination, whose associates,  $A_2$  and  $B_2$ , had been individually trained in Phase 1, and for the  $F_2-H_2$  discrimination, whose associates,  $E_2$  and  $G_2$ , had not been previously trained.

The learning curves in Figure 2 (left) suggest a facilitation of the  $X_1-Y_1$  discrimination, as well as an impairment of the  $X_2-Y_2$  discrimination, as compared with the novel element discrimination,  $W-Z$ . The difference between the  $X_1-Y_1$  discrimination and the novel  $W-Z$  discrimination, although in the correct direction, fell short of statistical significance [ $F(1,42) = 1.85$ ,  $MS_e = 0.13$ ,  $p = .18$ ]. The difference between the  $X_2-Y_2$  discrimination and the novel  $W-Z$  discrimination fell short of statistical significance as well [ $F(1,42) = 1.09$ ,  $MS_e = 0.18$ ]. Nonetheless, this tendency toward superior performance of the  $X_1-Y_1$  discrimination over the novel  $W-Z$  discrimination again suggests that learning of the  $X_1-Y_1$  discrimination was taking place in Phase 2, as is predicted by CH, but not by RW. Interestingly, the tendency for the  $X_2-Y_2$  discrimination to be impaired, relative to the novel  $W-Z$  discrimination, suggests that training of the unsolvable  $X_2-Y_2$  discrimination in Phase 2 might have had a deleterious effect on later learning of this discrimination. Perhaps the participants learned in Phase 2 that Cues  $X_2$  and  $Y_2$  were unrelated to the outcome, so that later learning in Phase 3 that  $X_2$  predicts the occurrence of the outcome and  $Y_2$  predicts the nonoccurrence of the outcome takes longer than learning of the novel discrimination.

Nevertheless, these nonsignificant trends involving the novel  $W-Z$  discrimination do not allow us to reach any definitive conclusions.

**Ratings.** Mean ratings of all 18 cues are shown in Figure 3. Because each of the cues—except  $E_1$ ,  $G_1$ ,  $E_2$ , and  $G_2$ —had been trained alone to asymptote, we were not expecting notably different ratings beyond extremely positive ratings for cues paired with the occurrence of the outcome and extremely negative ratings for cues paired with the nonoccurrence of the outcome. Indeed, mean ratings did largely reflect the contingencies that were given in training; however, we also observed that ratings were not equally high or equally low among the positive and the negative cues.

To further analyze these data in a more readily understandable manner, we again calculated a discrimination score for each of the discriminations illustrated in Phase 3. For comparison purposes, we also included the two discriminations that were learned in Phase 1.<sup>2</sup> In order to calculate the discrimination score ratings, we subtracted the rating for each cue paired with the nonoccurrence of the outcome from the rating for the corresponding cue paired with the occurrence of the outcome. Because our rating scale ranged from  $-100$  to  $+100$ , our discrimination scores varied along a 200-point scale. Mean scores near 200 represent excellent discrimination (i.e., a rating of 100 to the positive cue and a rating of  $-100$  to the corresponding negative cue), whereas mean scores near 0 represent no discrimination (similar ratings to the positive and the negative cues). Thus, we obtained seven discrimination scores:  $A_1-B_1$ ,  $A_2-B_2$ ,  $X_1-Y_1$ ,  $X_2-Y_2$ ,  $F_1-H_1$ ,  $F_2-H_2$ , and  $W-Z$ , which are displayed in Figure 4.

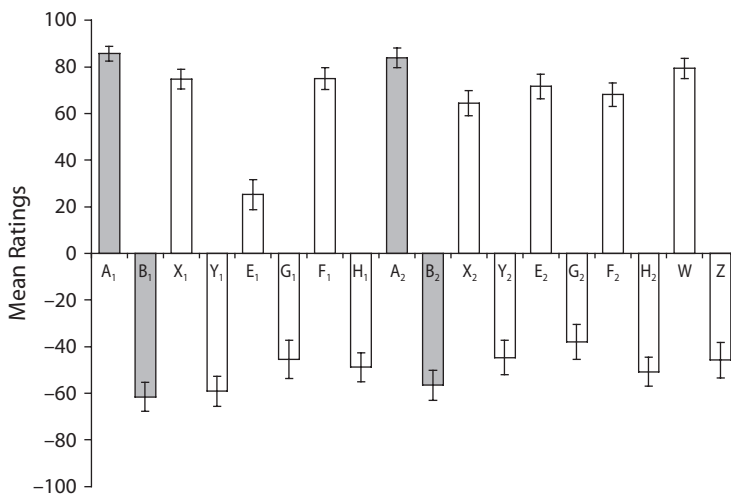
Surprisingly, even though all of these discriminations had been trained to asymptote, the derived ratings difference scores were not identical. Specifically, the  $X_1-Y_1$  discrimination score was higher than the  $X_2-Y_2$  discrimination score, with the novel discrimination score,  $W-Z$ , falling in between. This ordering neatly parallels the par-

ticipants' outcome predictions in Phase 3: Discrimination learning in the consistent condition was faster than that in the novel condition, with discrimination learning in the inconsistent condition being the slowest of the three. A one-way ANOVA of the ratings difference scores revealed a significant main effect [ $F(6,252) = 4.71$ ,  $MS_e = 1,564.56$ ]. Planned comparisons confirmed that the  $X_1-Y_1$  ratings discrimination was significantly higher than the  $X_2-Y_2$  ratings discrimination [ $F(1,42) = 8.83$ ,  $MS_e = 880.72$ ]. Thus, the advantage of the consistent over the inconsistent condition still remained on postdiscrimination ratings, again supporting the idea that the participants were effectively learning about the  $X_1-Y_1$  discrimination in Phase 2. None of the other target ratings discriminations differed reliably from one another.

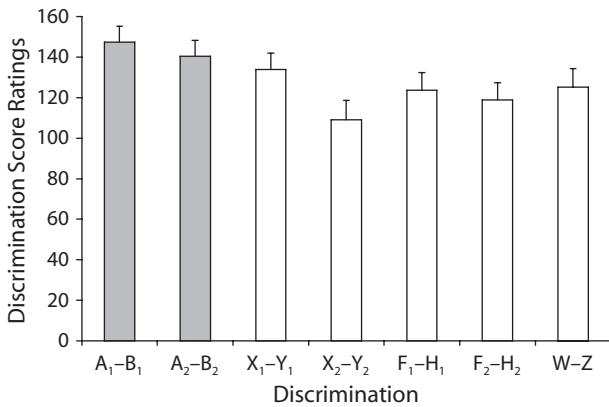
Overall, the results of Experiment 1 show that after learning an  $A_1+B_1-$  discrimination, the participants did learn about added redundant discriminative stimuli ( $X_1+/Y_1-$ ) that were presented in compound with the initial discriminative stimuli ( $A_1X_1+/B_1Y_1-$ ). The RW model predicts no  $X_1+/Y_1-$  learning, so that the associative values of  $X_1$  and  $Y_1$  should be the same after a solvable ( $X_1+/Y_1-$ ) discrimination or after an unsolvable ( $X_2\pm/Y_2\pm$ ) discrimination. Nevertheless, elemental discrimination training in Phase 3 showed that the participants' learning of  $X_1+/Y_1-$  proceeded more quickly than did their learning of  $X_2+/Y_2-$ , suggesting that learning about the added discriminative stimuli had taken place. In addition, mean postdiscrimination ratings difference scores disclosed an advantage of  $X_1-Y_1$  over  $X_2-Y_2$ . Hence, these results support a performance account, such as CH, but not a competitive-learning account, such as RW.

## EXPERIMENT 2

To further pursue the possible learning about the added Phase 2 discriminations, in our second experiment, we gave the participants the same Phase 1 and Phase 2 train-



**Figure 3.** Mean final ratings of all the cues in Experiment 1. The discriminations that were presented in Phase 1 are shown in gray. Error bars indicate the standard error of the mean.



**Figure 4.** Mean discrimination scores of the final ratings in Experiment 1. The discriminations that were presented in Phase 1 are shown in gray. Our target discriminations, presented in Phase 3, are shown in white. Error bars indicate the standard error of the mean.

ing as in Experiment 1. In Phase 3, however, the discriminations were *reversed* for the consistent condition (see Table 1). If the X<sub>1</sub>+ /Y<sub>1</sub>- discrimination was acquired in Phase 2, learning of the X<sub>1</sub>- /Y<sub>1</sub>+ discrimination should now proceed more *slowly* in the consistent than in the inconsistent condition (for which the unsolvable X<sub>2</sub>± /Y<sub>2</sub>± discrimination was given in Phase 2). However, if no learning about Cues X<sub>1</sub> and Y<sub>1</sub> occurred in Phase 2, learning of the Phase 3 discrimination should proceed similarly in both conditions.

**Method**

A total of 38 students at the University of Iowa received course credit for their voluntary participation. The apparatus, materials, and procedures were identical to those in Experiment 1, except for the reversal of the discriminations in Phase 3.

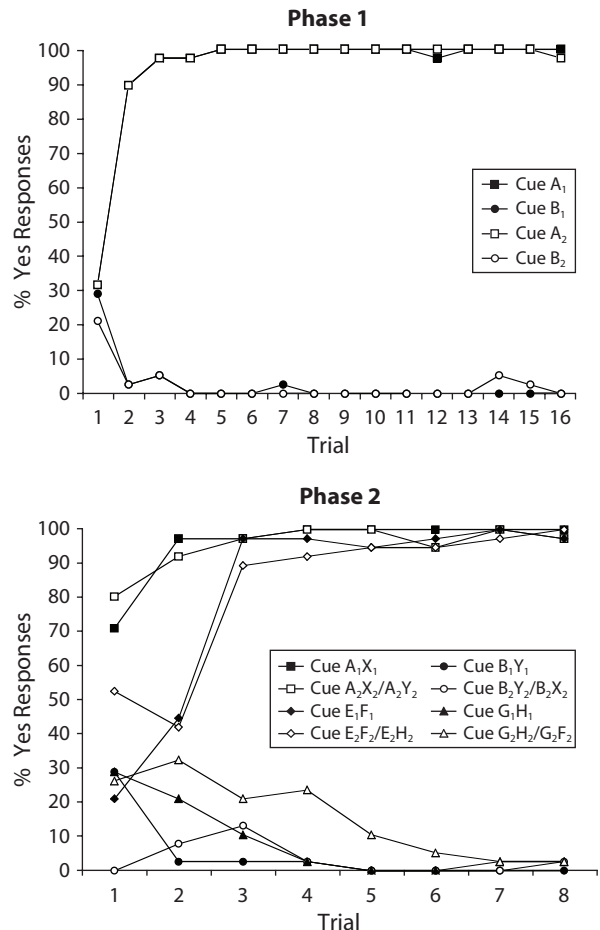
The experimental design is summarized in Table 1. After presenting A<sub>1</sub>X<sub>1</sub>+ /B<sub>1</sub>Y<sub>1</sub>- and E<sub>1</sub>F<sub>1</sub>+ /G<sub>1</sub>H<sub>1</sub>- discrimination training in Phase 2, we trained the participants in Phase 3 with X<sub>1</sub>- /Y<sub>1</sub>+ and F<sub>1</sub>- /G<sub>1</sub>+ discriminations, entailing a reversal of the Phase 2 reinforcement contingencies. For cues involved in the inconsistent condition, it makes no sense to speak about reversed discriminations, because no solvable discrimination was presented in Phase 2; for comparison purposes, we presented X<sub>2</sub>- /Y<sub>2</sub>+ and F<sub>2</sub>- /H<sub>2</sub>+ discriminations in Phase 3. A novel discrimination, W+ /Z-, was included as well.

**Results and Discussion**

**Training: Phases 1 and 2.** Acquisition behavior in the first and second phases of training is illustrated in Figure 5. As in Experiment 1, the participants very rapidly and virtually perfectly learned the A<sub>1</sub>+ , B<sub>1</sub>- and A<sub>2</sub>+ , B<sub>2</sub>- discriminations that were given in Phase 1 (see Figure 5, top). Beyond the second trial with each of the four cues, the participants exhibited asymptotic discriminative performance; the participants correctly predicted the occurrence of the outcome after Cues A<sub>1</sub> and A<sub>2</sub> and the nonoccurrence of the outcome after Cues B<sub>1</sub> and B<sub>2</sub>. This high level of discriminative performance to each of the four cues was maintained throughout the final 14 trials of Phase 1.

In Phase 2 (see Figure 5, bottom), although there was again a small decrement in discriminative performance on the first trial, the participants' accuracy was very high in the consistent condition: They correctly predicted the occurrence of the outcome after A<sub>1</sub>X<sub>1</sub> and the nonoccurrence of the outcome after B<sub>1</sub>Y<sub>1</sub>. As well, in the inconsistent condition, the participants accurately predicted the occurrence of the outcome after A<sub>2</sub>X<sub>2</sub> and A<sub>2</sub>Y<sub>2</sub> trials and the nonoccurrence of the outcome after B<sub>2</sub>X<sub>2</sub> and B<sub>2</sub>Y<sub>2</sub> trials. There were again no clear differences in discriminative performance between the consistent and the inconsistent conditions.

The two compound control discriminations generally emerged more slowly than the discriminations involving only elements that had been presented in Phase 1 (Figure 5, bottom). In addition, E<sub>1</sub>F<sub>1</sub>+ /G<sub>1</sub>H<sub>1</sub>- learning was faster than E<sub>2</sub>F<sub>2</sub>+ , E<sub>2</sub>H<sub>2</sub>+ /G<sub>2</sub>F<sub>2</sub>- , G<sub>2</sub>H<sub>2</sub>- learning; this disparity in discrimination learning between the consistent and the inconsistent conditions did not occur in the case of the A<sub>1</sub>X<sub>1</sub>+ /B<sub>1</sub>Y<sub>1</sub>- and A<sub>2</sub>X<sub>2</sub>+ , A<sub>2</sub>Y<sub>2</sub>+ /B<sub>2</sub>X<sub>2</sub>- , B<sub>2</sub>Y<sub>2</sub>- discriminations in Phase 1. Nonetheless, by the



**Figure 5.** Percentages of predictions of the occurrence of the outcome after each of the cues in Phase 1 (top) and Phase 2 (bottom) of Experiment 2. Error bars indicate the standard error of the mean.

fifth trial, the participants were able to accurately predict the occurrence and nonoccurrence of the outcome after each of the compound cues. The results of Phases 1 and 2 in this experiment paralleled the results of Experiment 1, as should be the case, because training up to this point was the same.

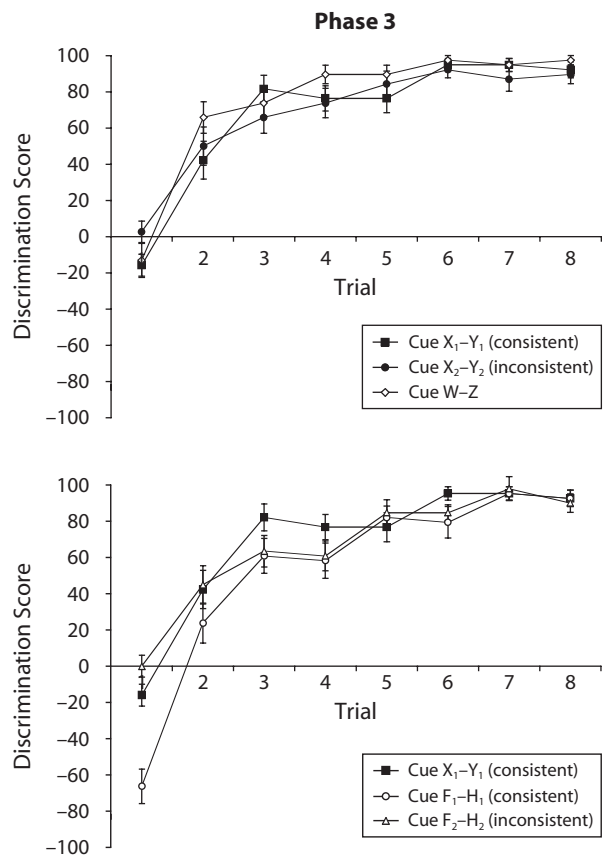
**Training: Phase 3.** To analyze learning of the different discriminations in Phase 3, the key learning phase, we again calculated discrimination scores for each of the discriminations on every trial. From the prediction for each cue paired with the occurrence of the outcome, we subtracted the prediction for the corresponding cue paired with the nonoccurrence of the outcome. We obtained five different discrimination scores in Phase 3:  $X_1 - Y_1$ ,  $X_2 - Y_2$ ,  $F_1 - H_1$ ,  $F_2 - H_2$ , and  $W - Z$ .

The development of the consistent discrimination  $X_1 - Y_1$ , as compared with the inconsistent and novel discriminations, is displayed in Figure 6 (top); the development of the same consistent discrimination  $X_1 - Y_1$  compared with the control discriminations without Phase 1 training is displayed in Figure 6 (bottom). Here, we expected that, if learning occurred in Phase 2, the  $X_1 - Y_1$  discrimination would be impaired in Phase 3. Indeed, it appears that the  $X_1 - Y_1$  discrimination was generally worse than the novel  $W - Z$  discrimination, but it was similar to the  $X_2 - Y_2$  discrimination (Figure 6, top). As well, it appears that the other reversed discrimination,  $F_1 - H_1$ , was worse than the  $X_1 - Y_1$  discrimination (Figure 6, bottom), suggesting better learning of the *overshadowed*  $F_1 - H_1$  discrimination than of the *blocked*  $X_1 - Y_1$  discrimination in Phase 2.

A 5 (discrimination:  $X_1 - Y_1$ ,  $X_2 - Y_2$ ,  $F_1 - H_1$ ,  $F_2 - H_2$ , or  $W - Z$ )  $\times$  7 (trial: 2–8) ANOVA of Phase 3 discrimination scores revealed significant main effects of discrimination [ $F(4,148) = 3.44$ ,  $MS_e = 0.30$ ] and trial [ $F(6,222) = 31.71$ ,  $MS_e = 0.17$ ]. The discrimination  $\times$  trial interaction did not reach statistical significance [ $F(24,888) = 1.38$ ,  $MS_e = 0.13$ ,  $p = .10$ ].

Within this ANOVA, linear trend analyses confirmed that the  $X_1 - Y_1$  discrimination was learned more slowly than was the novel  $W - Z$  discrimination [ $F(1,37) = 4.20$ ,  $MS_e = 0.16$ ], suggesting that the participants did effectively learn the  $X_1+$  and  $Y_1-$  contingencies in Phase 2 and found it difficult to learn the opposite contingencies; this result supports the CH account. There was no difference between the  $X_1 - Y_1$  and the  $X_2 - Y_2$  discriminations [ $F(1,37) = 0.20$ ,  $MS_e = 0.33$ ], suggesting that Phase 2 training slowed both discriminations. In fact, the  $X_2 - Y_2$  discrimination was also learned more slowly than was the novel  $W - Z$  discrimination [ $F(1,37) = 5.31$ ,  $MS_e = 0.22$ ]. In the case of the  $X_1 - Y_1$  discrimination, the impairment presumably occurred because the participants had to reverse the contingencies that they had learned in Phase 2; in the case of the  $X_2 - Y_2$  discrimination, it might be that the participants learned in Phase 2 that  $X_2$  and  $Y_2$  were uncorrelated with outcome occurrence and nonoccurrence, also leading to slower discrimination learning involving those stimuli.

Although RW cannot explain the  $X_2 - Y_2$  impairment, CH potentially can. As we described above, according to



**Figure 6.** Percentages of predictions of the occurrence of the outcome in Phase 3 of Experiment 2. Discrimination scores are plotted. **Top:** The consistent discrimination  $X_1 - Y_1$  is compared with the inconsistent and novel discriminations. **Bottom:** The consistent discrimination  $X_1 - Y_1$  is compared with the control discriminations without training in Phase 1. Error bars indicate the standard error of the mean.

CH, both Cues  $X_2$  and  $Y_2$  should acquire some positive strength in Phase 2, so that at the beginning of Phase 3, both Cues  $X_2$  and  $Y_2$  should start with a similar positive value; then, learning to discriminate them should take longer than learning a novel discrimination. Nonetheless, although the  $X_2 - Y_2$  discrimination should be impaired, as compared with a novel discrimination, it should still be superior to the  $X_1 - Y_1$  discrimination. At the beginning of Phase 3,  $X_1$  should start with very high associative strength, and  $Y_1$  should start with very low associative strength, whereas both  $X_2$  and  $Y_2$  should start with similar and slightly positive associative strength. According to CH, learning the reversed positive and negative contingencies ( $X_1 - Y_1$  discrimination) should take longer than learning the positive and negative contingencies departing from an intermediate value ( $X_2 - Y_2$  discrimination); but we did not observe this difference.

Linear trend analyses also disclosed that the  $F_1 - H_1$  discrimination in the consistent condition and the  $F_2 - H_2$  discrimination in the inconsistent condition did not differ from each other [ $F(1,37) = 0.82$ ] and that both discriminations were slowed in relation to the novel  $W - Z$  discrim-

ination [ $F(1,37) = 11.35$ ,  $MS_e = 0.33$ , and  $F(1,37) = 7.6$ ,  $MS_e = 0.25$ , respectively]. Thus, the performance for the  $X_1-Y_1$  discrimination in relation to the  $X_2-Y_2$  discrimination was equivalent to the performance for the  $F_1-H_1$  discrimination in relation to the  $F_2-H_2$  discrimination, suggesting equivalent learning in Phase 2, regardless of whether or not their associates had been trained alone in Phase 1.

Yet we also found that the  $X_1-Y_1$  discrimination was not as impaired as the  $F_1-H_1$  discrimination [ $F(1,37) = 4.44$ ,  $MS_e = 0.16$ ], suggesting that the associations of  $F_1$  and  $H_1$  with the contingencies of reinforcement were better learned in Phase 2 than were the associations of  $X_1$  and  $Y_1$  with the same contingencies of reinforcement. This disparity between  $X_1-Y_1$  and  $F_1-H_1$  suggests that some blocking of the added  $X_1-Y_1$  discrimination might have taken place in Phase 2. This disparity provides evidence that supports RW, but not CH, because CH predicts no difference in learning these discriminations. We should note that this is the only hint in all of our data that suggests that some competition occurred during the acquisition phase.

**Ratings.** Mean ratings of all 18 cues are shown in Figure 7. Ratings generally reflected the contingencies that were presented in training. To analyze these ratings in a simpler fashion, we calculated a discrimination score for each of the discriminations presented in Phase 3, just as in Experiment 1. Again, we included the discriminations presented in Phase 1 for comparative purposes. Thus, we obtained seven discrimination scores:  $A_1-B_1$ ,  $A_2-B_2$ ,  $X_1-Y_1$ ,  $X_2-Y_2$ ,  $F_1-H_1$ ,  $F_2-H_2$ , and  $W-Z$ , which are displayed in Figure 8.

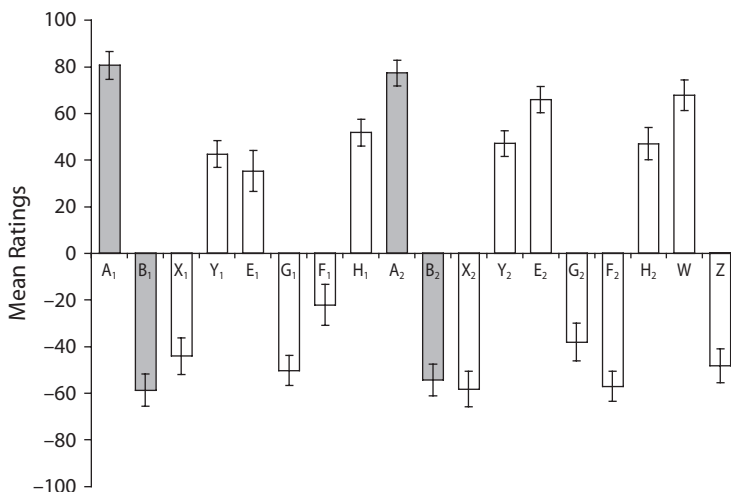
As in Experiment 1, although all of these discriminations had been trained to asymptote, the ratings difference scores were not identical. Especially interesting was the fact that the  $X_1-Y_1$  discrimination score was now lower than either the  $X_2-Y_2$  score or the  $W-Z$  score. No difference between the  $X_1-Y_1$  and  $X_2-Y_2$  discriminations had materialized in Phase 3 training: Both were equally

worse than the novel  $W-Z$  discrimination. Here, however, the ratings difference scores revealed that the  $X_1-Y_1$  discrimination was even more strongly impaired than was the  $X_2-Y_2$  discrimination.

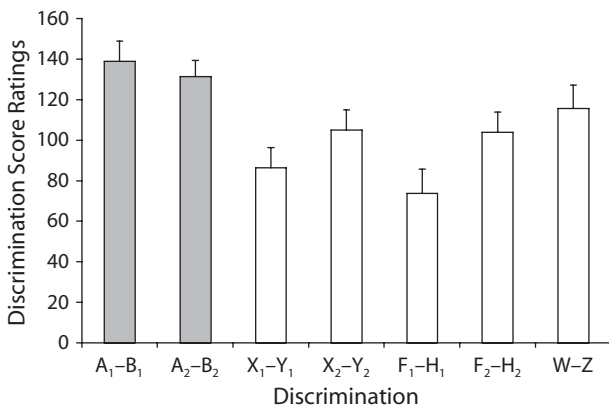
A one-way ANOVA of the ratings difference scores revealed a significant main effect [ $F(6,222) = 7.54$ ,  $MS_e = 2,719.79$ ]. Planned comparisons showed that the  $X_1-Y_1$  ratings discrimination was lower than the  $W-Z$  ratings discrimination [ $F(1,37) = 11.44$ ,  $MS_e = 1,411.12$ ], confirming that the impairment in Phase 3 persisted to the rating phase. The  $X_1-Y_1$  ratings discrimination was also lower than the  $X_2-Y_2$  ratings discrimination [ $F(1,37) = 3.84$ ,  $MS_e = 1,706.42$ ]. In Phase 3, the acquisition of the  $X_1-Y_1$  and  $X_2-Y_2$  discriminations had proceeded similarly; however, ratings showed that consistent training with the opposite contingencies in Phase 2 (as was the case for the  $X_1-Y_1$  discrimination) was more harmful than was inconsistent training (as was the case for the  $X_2-Y_2$  discrimination). This was the prediction of the CH that did not materialize in Phase 3: Learning the reversed positive and negative contingencies ( $X_1-Y_1$  discrimination) should have taken longer than learning the positive and negative contingencies when each cue's strength began at an intermediate value ( $X_2-Y_2$  discrimination). Although that difference did not appear in the learning phase, it did surface in later ratings.

As well, ratings of the  $F_1-H_1$  discrimination were lower than ratings of the  $F_2-H_2$  discrimination [ $F(1,37) = 5.85$ ,  $MS_e = 2,951.28$ ] and lower than ratings of the  $W-Z$  discrimination [ $F(1,37) = 7.51$ ,  $MS_e = 4,453.66$ ]. These impairments are equivalent to the  $X_1-Y_1$  ratings impairments; this equivalence was predicted by CH, but not by RW. The difference between the  $X_1-Y_1$  and the  $F_1-H_1$  discriminations in Phase 3 training, which suggested some blocking, did not appear in the ratings phase [ $F(1,37) = 0.87$ ].

Overall, the results of Experiment 2 reveal that after training an  $A_1+B_1-$  discrimination to asymptote in Phase 1, the participants were able to learn about an added



**Figure 7.** Mean final ratings of all the cues in Experiment 2. The discriminations that were presented in Phase 1 are shown in gray. Error bars indicate the standard error of the mean.



**Figure 8.** Mean discrimination scores of the final ratings in Experiment 2. The discriminations that were presented in Phase 1 are shown in dark gray. Our target discriminations, presented in Phase 3, are shown in white. Error bars indicate the standard error of the mean.

$X_1 + Y_1 -$  discrimination in Phase 2; when the added discrimination was reversed in Phase 3, reversal learning was impaired, relative to a novel discrimination, although not as impaired as an *overshadowed* discrimination. The RW model predicts no learning of  $X_1 + Y_1 -$  in Phase 2  $A_1 X_1 + B_1 Y_1 -$  training, so there should be no interference in Phase 3  $X_1 - Y_1 +$  learning. However, elemental discrimination training in Phase 3 disclosed that the participants' learning of  $X_1 - Y_1 +$  proceeded more slowly than did their learning of a novel discrimination,  $W - Z +$ , suggesting that learning of the added discrimination had, in fact, taken place. Nonetheless, some of the blocking predicted by RW might have taken place, because after Phase 2  $E_1 F_1 + / G_1 H_1 -$  training, the participants' learning of  $F_1 - / H_1 +$  proceeded even more slowly than their learning of  $X_1 - / Y_1 +$ . Although these results accord better with a performance account, such as CH, than with a competitive-learning account, such as RW, the difference between  $X_1 - / Y_1 +$  and  $F_1 - / H_1 +$  is better explained by RW.

## GENERAL DISCUSSION

Our participants first solved a single discrimination, and when their performance had clearly reached asymptote, a new redundant discrimination was added. In Experiment 1, when this added discrimination was later trained alone, we observed substantial *savings*, as compared with nondiscrimination or novel discrimination controls. In Experiment 2, when we later reversed the contingencies of the added discrimination, we observed a significant performance *impairment*, as compared with a novel discrimination. Moreover, in both experiments, after training was complete, the participants' causal judgments of the added discrimination still reflected these differences. Thus, we conclude that there was robust learning about an added discrimination even when it did not provide any new information about the experimental contingencies. These experiments add to prior studies suggesting that performance, rather than acquisition, deficits underlie the

blocking effect (Blaisdell et al., 1999; Denniston et al., 2003).

## The Role of Within-Compound Associations

According to CH, responding to a target cue is determined not only by the association between that stimulus and the outcome, but also by the association between other stimuli (comparator cues) and the same outcome. A comparator cue will effectively compete with a target cue for the response as long as it has a strong association with the outcome. If the comparator cue is further trained without being followed by the outcome, the comparator–outcome association will be weakened, and the comparator cue will no longer effectively compete with the target cue for the expression of the response. Not only does the comparator–outcome association have to be strong for the comparator cue to effectively compete with the target cue, but also there has to be a firm *within-compound* association between the target cue and the comparator cue. If either the comparator–outcome association or the target-cue–comparator association is weak or absent, the target-cue–outcome association will be able to be expressed, and a conditioned response to the target cue will be observed.

Consider further the role of within-compound associations in our experiments. After Phase 2, the within-compound associations of  $A_1$  with  $X_1$  and  $B_1$  with  $Y_1$  in the consistent condition were likely to be stronger than the within-compound associations of  $A_2$  with  $X_2$  and  $B_2$  with  $Y_2$  in the inconsistent condition; in the consistent condition,  $X_1$  was always presented along with  $A_1$ , and  $B_1$  was always presented along with  $Y_1$ , whereas in the inconsistent condition,  $X_2$  was paired on half of the trials with  $A_2$  and on half of the trials with  $B_2$ , and  $Y_2$  was paired on half of the trials with  $A_2$  and on half of the trials with  $B_2$ . These varied pairings in the inconsistent condition ought to have slowed the development of robust within-compound associations. The existence of strong within-compound associations in the consistent condition should have allowed comparators  $A_1$  and  $B_1$  to interfere with  $X_1$  and  $Y_1$  at the beginning of Phase 3; this interference should not have taken place or should have been milder in the inconsistent condition, in which the within-compound associations were weaker. This possibly stronger interference of the comparator cues in the consistent condition might impair the expression of the  $X_1 - Y_1$  discrimination in Phase 3. So, even if there were learning of the  $X_1 - Y_1$  discrimination in Phase 2, that learning should not become evident because of the influence of the comparator cues. The disparity in the strength of the within-compound associations in the consistent and the inconsistent conditions should never reverse the predictions of CH; if anything, that disparity should produce weaker effects. Had it been possible to have within-compound associations of similar strengths, we might have been able to report even stronger effects.

## Configural Learning

Our consideration above of within-compound associations implies that each of the elements of the compound retains its particular properties: An association is formed

between or among *individual* cues, but each of the cues *independently* enters into an association with the outcome. However, it might be the case that our participants learned cue *configurations*, as suggested by Pearce's (1987, 1994, 2002) theoretical proposals. According to this view, when cues are presented in compound (e.g., AX), the entire pattern or configuration will become associated with the outcome. When one of the elements is later presented alone (e.g., X), the response will be determined by its degree of similarity to the trained configural *unit*. In Phase 2 of our Experiment 1,  $A_1X_1+$  and  $B_1Y_1-$  were presented in the consistent condition, whereas  $A_2X_2+$ ,  $A_2Y_2+$ ,  $B_2X_2-$ , and  $B_2Y_2-$  were presented in the inconsistent condition. In Phase 3, the presentation of  $X_1$  should excite configural unit  $A_1X_1$ , and half of its positive associative strength should generalize to  $X_1$ . In a parallel fashion, the presentation of  $Y_1$  should excite configural unit  $B_1Y_1$ , and half of its negative associative strength should generalize to  $Y_1$ . Therefore, the  $X_1-Y_1$  discrimination should start with some advantage, due to the associative strengths generalized from the configural units that were trained in Phase 2.

The situation is different in the inconsistent condition. In this case, the presentation of  $X_2$  in Phase 3 should excite configural unit  $A_2X_2$  and configural unit  $B_2X_2$ ;  $A_2X_2$  had been paired with the outcome, whereas  $B_2X_2$  had been paired with the absence of the outcome. Cue  $X_2$  should then activate two configural units that carry opposite associative value, so they should cancel one another. The same holds true for Cue  $Y_2$ . The presentation of  $Y_2$  in Phase 3 should excite configural unit  $A_2Y_2$  and configural unit  $B_2Y_2$ ; because  $A_2Y_2$  had been paired with the outcome and  $B_2Y_2$  had been paired with the absence of the outcome, Cue  $Y_2$  should activate two configural units of opposite value, so that they will again cancel one another. Hence, the  $X_2-Y_2$  discrimination will start from a lower level than will the  $X_1-Y_1$  discrimination because of differential generalization from Phase 2 training, not because the participants learned about the solvable  $X_1-Y_1$  discrimination and the nonsolvable  $X_2-Y_2$  discrimination in Phase 2.

Although Pearce's (1987, 1994, 2002) configural theory can explain the difference between the  $X_1-Y_1$  and the  $X_2-Y_2$  discriminations in Experiment 1, it cannot explain the fact that, after the first trial of Phase 3, there was no clear difference between the  $X_1-Y_1$  and the  $F_1-H_1$  discriminations. According to Pearce, after Phase 2, the  $E_1F_1$  and  $G_1H_1$  configural units should have maximum positive and negative strength, respectively; so later in Phase 3, half of the positive strength will generalize to  $F_1$ , and half of the negative strength will generalize to  $H_1$ . However, when  $A_1X_1$  and  $B_1Y_1$  are presented in Phase 2, the occurrence and nonoccurrence of the outcome is already partially predicted by  $A_1$  and  $B_1$ , respectively. Configural units  $A_1X_1$  and  $B_1Y_1$  in this case can gain only half of the associative strength that they could have gained if  $A_1$  and  $B_1$  had not been previously trained. So, the asymptotic associative strength of compounds  $A_1X_1$  and  $B_1Y_1$  at the end of Phase 2 will be half of the associative strength of compounds  $E_1F_1$  and  $G_1H_1$ . In Phase 3, when  $X_1$  and

$Y_1$  are later presented alone, they will activate the corresponding configural units to half their level, which is half the value that  $F_1$  and  $H_1$  are able to activate. In this way, Pearce's configural theory can explain the blocking effect (the difference between the  $X_1-Y_1$  and the  $F_1-H_1$  discriminations on the first trial of Phase 3), but it cannot explain the fact that, by the second trial of Phase 3, the two discriminations are at the same high level. Given that the salience of our cues is equal—because they were fully counterbalanced—learning should proceed at a similar rate for both discriminations.

Although we did not find any disparity between learning of the  $X_1-Y_1$  and the  $F_1-H_1$  discriminations in Experiment 1, we did find a disparity in Experiment 2. As we just described, Pearce's (1987, 1994, 2002) configural theory could readily explain it. But Pearce's theory also predicts a difference between the  $X_1-Y_1$  and the  $X_2-Y_2$  discriminations that we did not observe. After training of  $A_1X_1+$  and  $B_1Y_1-$  in Phase 2, the presentation of  $X_1$  at the beginning of Phase 3 should excite configural unit  $A_1X_1$ , and half of its positive associative strength should generalize to  $X_1$ . In a parallel fashion, the presentation of  $Y_1$  should excite configural unit  $B_1Y_1$ , and half of its negative associative strength should generalize to  $Y_1$ . So, the  $X_1-Y_1$  discrimination should start with some *disadvantage*, because in Phase 3, the opposite contingencies had to be learned. On the other hand, after training of  $A_2X_2+$ ,  $A_2Y_2+$ ,  $B_2X_2-$ , and  $B_2Y_2-$  in Phase 2, the presentation of  $X_2$  in Phase 3 should excite configural unit  $A_2X_2$  and configural unit  $B_2X_2$ . Because  $A_2X_2$  had been paired with the outcome and  $B_2X_2$  had been paired with the absence of the outcome,  $X_2$  should activate two configural units that carry opposite values, so they should cancel one another. The same holds true for  $Y_2$ . The presentation of  $Y_2$  in Phase 3 should excite configural unit  $A_2Y_2$  and configural unit  $B_2Y_2$ ; because  $A_2Y_2$  had been paired with the outcome and  $B_2Y_2$  had been paired with the absence of the outcome,  $Y_2$  should activate two configural units of opposite value, so that they will again cancel one another. So, the  $X_2-Y_2$  discrimination should start from a *neutral* level, as compared with the *opposite* level of the  $X_1-Y_1$  discrimination; but we did not observe this disparity.

### When Did Learning Take Place?

Kamin (1968) studied the rate of acquisition of a conditioned response to blocked Cue X, as compared with a novel cue. He found that although there was no difference on the first training trial, more conditioned responding to blocked Cue X than to the novel cue occurred on later trials; Cue X learning showed some small savings in relation to a novel cue—just as in our study. On the other hand, Kamin also found that, after A+ training, the conditioned response on the first trial of AX+ training was not as high as it had been on the last trial of A+ alone training. The introduction of Cue X caused some disruption of conditioned responding, as was also the case in our study when responding to an excitatory stimulus was decreased (and when responding to an inhibitory stimulus was increased) (see Figure 1, bottom; Figure 5, bottom).

Combining these two findings, Kamin (1968) conjectured that some learning about Cue X might occur on the first compound trial; that learning could be responsible for the savings in the later acquisition of Cue X. When Kamin compared the rate of acquisition of a blocked cue given the standard procedure with several compound training trials with the rate of acquisition of another cue given only one compound training trial, he found no differences; so he concluded that it had to be on the first AX+ trial, and only on that trial, that Cue X acquired associative strength.

Consistent with Kamin's (1968) argument, Mackintosh (1975a) found no evidence of blocking on the first compound trial in a blocking procedure (but see Balaz, Kaspro, & Miller, 1982, for blocking with just one compound trial). According to Mackintosh (1975a), blocking should develop in later trials if organisms learn to *ignore* the added cue, because that cue does not predict any *change* in the contingencies of reinforcement. Blocking may be due not to competition for associative strength (as RW contends) or even for attention, but to the fact that attention is not summoned to cues that provide no *new* information.

Could the advantage of the  $X_1 - Y_1$  discrimination in our study have been due to learning that occurred on just that first compound trial? We compared the last trial of A and B in Phase 1 with the first trial of AX and BY in Phase 2, respectively (see Figures 1 and 5). In both experiments, mean predictions for A at the end of Phase 1 were 100 and mean predictions for B were 0. When X was added, mean predictions for AX, although still high, decreased ( $M = 81.3$  and  $71.0$ , for Experiments 1 and 2, respectively), and the differences from the last trial of Phase 1 were significant [ $t(42) = 3.09$ ,  $p = .003$ , and  $t(37) = 3.88$ ,  $p < .001$ , for Experiments 1 and 2, respectively]. When Y was added, mean predictions for BY, although still low, increased ( $M = 20.9$  and  $28.9$ , for Experiments 1 and 2, respectively), and the differences from the last trial of Phase 1 were significant [ $t(42) = -3.33$ ,  $p = .001$ , and  $t(37) = -3.88$ ,  $p < .001$ , for Experiments 1 and 2, respectively]. Hence, at the beginning of Phase 2, there may have been a discrepancy between the outcome that the participants expected and the actual outcome. Perhaps this discrepancy permits some learning of the added  $X_1 - Y_1$  discrimination, so that the facilitation that we observed for  $X_1 - Y_1$  in Phase 3 was due to this first-trial learning.

RW has no principled mechanism that reduces the outcome expectation due to changes in the environmental situation, making it impossible for RW to explain our data. But Mackintosh's (1975b) attentional theory could embrace our results, because it assumes that organisms normally attend to novel cues; so, organisms might condition to an added cue on the first trial on which it is presented (Pearce & Hall, 1980, made the same prediction). On subsequent trials, when organisms appreciate that the added cue does not provide any new information, no more learning should occur. To demonstrate that the facilitation that we observed for  $X_1 - Y_1$  in Phase 3 was due to some associative learning occurring on just the first trial would require more empirical data than we presently have.

## Learning Versus Performance

[Although] the distinction between learning and performance is in fact often used, explicitly or implicitly, in the formulation of the experimental problems, there has been relatively little systematic concern with the determiners of performance as distinguished from the conditions of learning. (Postman, 1968, p. 556)

Postman's (1968) cogent complaint was eclipsed in the 1970s by the success of error correction learning models such as those of Rescorla and Wagner (1972), Mackintosh (1975a), and Pearce and Hall (1980). Low levels of responding to a cue that had been paired with an outcome were conceptualized as learning deficits. The comparator hypothesis (Denniston et al., 2001; Miller & Matzel, 1988) revived the learning versus performance debate by postulating that response deficits might be due not to acquisition failures, but to performance processes. Organisms would be able to store information about events occurring together and later use that information when the appropriate time came.

Overall, our results support the view that organisms do, indeed, learn about cues that do not signal any change in the environment—information that is redundant and not immediately useful. This finding poses a serious problem for RW, but it supports CH. It seems that, at least in our case, contiguity between cue and outcome (or its absence) is sufficient for learning to occur. But not all of our data can be explained by the CH; there was some hint of competitive learning in our results as well (the disparity between the  $X_1 - Y_1$  and  $F_1 - H_1$  discriminations in Experiment 2). And there are still other phenomena that are better accommodated by RW or different models that focus on learning—for example, the gradual acquisition and extinction of a cue–outcome association and trial order effects (see Denniston et al., 2001). The debate is far from over. The learning versus performance distinction is again a lively theoretical contest in the realms of both animal and human behavior. The experimental pursuit of this distinction should continue to stimulate interesting research and theory.

## AUTHOR NOTE

We thank Ralph R. Miller for many fertile and insightful discussions about our study. We also thank Oskar Pineño for his helpful comments on a prior version of the manuscript. Correspondence concerning this article should be addressed to L. Castro, Department of Psychology, University of Iowa, E11 Seashore Hall, Iowa City, IA 52242 (e-mail: leyre-castroruiz@uiowa.edu).

## REFERENCES

- ALLAN, L. G. (1993). Human contingency judgments: Rule based or associative? *Psychological Bulletin*, *114*, 435-448.
- BALAZ, M. A., KASPRO, W. J., & MILLER, R. R. (1982). Blocking with a single compound trial. *Animal Learning & Behavior*, *10*, 271-276.
- BLAISDELL, A. P., GUNTHER, L. M., & MILLER, R. R. (1999). Recovery from blocking achieved by extinguishing the blocking CS. *Animal Learning & Behavior*, *27*, 63-76.
- DENNISTON, J. C., SAVASTANO, H. I., BLAISDELL, A. P., & MILLER, R. R. (2003). Cue competition as a retrieval deficit. *Learning & Motivation*, *34*, 1-31.

- DENNISTON, J. C., SAVASTANO, H. I., & MILLER, R. R. (2001). The extended comparator hypothesis: Learning by contiguity, responding by relative strength. In R. R. Mowrer & S. B. Klein (Eds.), *Handbook of contemporary learning theories* (pp. 65-117). Hillsdale, NJ: Erlbaum.
- DICKINSON, A., & BURKE, J. (1996). Within-compound associations mediate the retrospective reevaluation of causality judgements. *Quarterly Journal of Experimental Psychology*, **49B**, 60-80.
- JOHNSON, D. F., & CUMMING, W. W. (1968). Some determiners of attention. *Journal of the Experimental Analysis of Behavior*, **11**, 157-166.
- KAMIN, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior: Aversive stimulation* (pp. 9-31). Miami: University of Miami Press.
- MACKINTOSH, N. J. (1975a). Blocking of conditioned suppression: Role of the first compound trial. *Journal of Experimental Psychology: Animal Behavior Processes*, **1**, 335-345.
- MACKINTOSH, N. J. (1975b). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review*, **82**, 276-298.
- McLAREN, I. P. L., & MACKINTOSH, N. J. (2000). An elemental model of associative learning: I. Latent inhibition and perceptual learning. *Animal Learning & Behavior*, **28**, 211-246.
- McLAREN, I. P. L., & MACKINTOSH, N. J. (2002). Associative learning and elemental representation: II. Generalization and discrimination. *Animal Learning & Behavior*, **30**, 177-200.
- MILLER, R. R., & MATZEL, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 51-92). San Diego: Academic Press.
- PEARCE, J. M. (1987). A model for stimulus generalization in Pavlovian conditioning. *Psychological Review*, **94**, 61-73.
- PEARCE, J. M. (1994). Similarity and discrimination: A selective review and a connectionist model. *Psychological Review*, **101**, 587-607.
- PEARCE, J. M. (2002). Evaluation and development of a connectionist theory of configural learning. *Animal Learning & Behavior*, **30**, 73-95.
- 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.
- POSTMAN, L. (1968). Association and performance in the analysis of verbal learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 550-571). Englewood Cliffs, NJ: Prentice-Hall.
- RAUHUT, A. S., MCPHEE, J. E., & AYRES, J. J. B. (1999). Blocked and overshadowed stimuli are weakened in their ability to serve as blockers and second-order reinforcers in Pavlovian fear conditioning. *Journal of Experimental Psychology: Animal Behavior Processes*, **25**, 45-67.
- RESCORLA, R. A., & WAGNER, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.
- SHANKS, D. R. (1995). *The psychology of associative learning*. Cambridge: Cambridge University Press.
- VAN HAMME, L. J., & WASSERMAN, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning & Motivation*, **25**, 127-151.
- WILLIAMS, B. A. (1996). Evidence that blocking is due to associative deficit: Blocking history affects the degree of subsequent associative competition. *Psychonomic Bulletin & Review*, **3**, 71-74.
- YOUNG, M. E., & WASSERMAN, E. A. (2004). Theories of learning. In K. Lamberts & R. Goldstone (Eds.), *Handbook of cognition* (pp. 161-182). Thousand Oaks, CA: Sage.

## NOTES

1. We did not include Trial 1 of Phase 3 in this and later analyses because prediction responses were made *before* participants saw the actual outcome, so this first trial could not reveal any learning effect.
2. Note that Cues A<sub>1</sub>, B<sub>1</sub>, A<sub>2</sub>, and B<sub>2</sub> were presented *alone* twice as often as any of the other cues, so the discriminations involving these cues are expected to be superior to the rest of the discriminations. Also, note that Cues E<sub>1</sub>, G<sub>1</sub>, E<sub>2</sub>, and G<sub>2</sub> will not be included in these analyses, because they were never presented alone.

## APPENDIX

### Instructions for the Experiments

We would like you to imagine that you are an allergist, that is, you are someone who tries to discover the cause of allergic reactions in people. You have just been presented with a new patient, "Mr. X," who suffers from allergic reactions when he eats. In an attempt to discover which foods cause him to have allergic reactions, you arrange for him to eat various foods for a meal on each day and you observe if he has an allergic reaction or not. The results of the daily allergy tests will be shown to you on a series of screens. You will see a separate screen for each day of the allergy test. On each screen you will be told what Mr. X ate that day and if there was an allergic reaction. Sometimes you will be shown the name of one food and sometimes two. Please pay attention to the different foods names and remember that your task as an allergist is to determine which food or foods are causing an allergic reaction.

After seeing each day's foods, you will be asked to predict whether or not each meal caused an allergic reaction in your patient. Simply click YES if you believe that your patient will suffer an allergic reaction and click NO if you think your patient will not suffer an allergic reaction. After you make your prediction, the computer will inform you whether or not Mr. X actually suffered from a reaction. Obviously, at first you will have to guess because you will not know anything about your patient, but hopefully you will begin to learn which foods cause him an allergic reaction and which do not. You might view this experiment as a game and try to score as many points (correct predictions) as you can. You will see the number of correct predictions that you have made near the bottom of the screen during the daily allergy tests.

Later in the experiment, you will be asked to rate to what extent each of the foods affects the likelihood of an allergic reaction in Mr. X, your patient, based on the information you have seen so far.